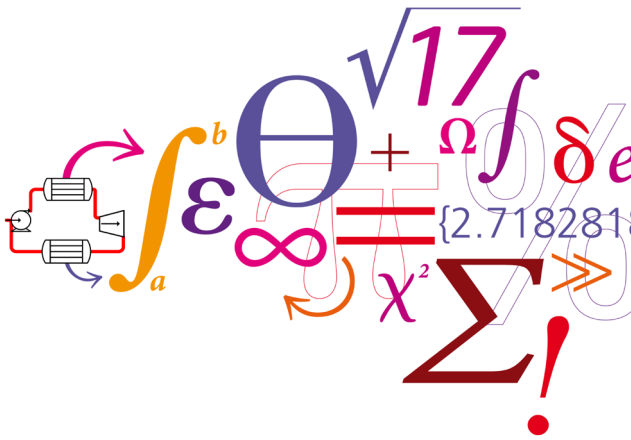
Property Uncertainty Analysis and Methods for Optimal Working Fluids of Thermodynamic Cycles

Frutiger, Jerome

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Property uncertainty analysis and methods for optimal working fluids of thermodynamic cycles



Jérôme Frutiger

PhD Thesis

May 2017

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Technical University of Denmark

PhD Thesis

Property uncertainty analysis and methods for optimal working fluids of thermodynamic cycles

Jérôme Frutiger

2017-05-31

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Technical University of Denmark
Anker Engelundsvej 1
Building 101A
DK-2800, Kgs. Lyngby
Denmark
CVR-nr. 30 06 09 46
Phone: (+45) 45 25 25 25
Email: dtu@dtu.dk
www.dtu.dk

Printed by STEP

Summary

There is an increasing interest in recovering industrial waste heat at low temperatures (70-250°C). Thermodynamic cycles, such as heat pumps or organic Rankine cycles, can recover this heat and transfer it to other process streams or convert it into electricity. The working fluid, circulating around the cycle, is vital for the performance of the cycle. Computational modelling of working fluid properties and cycle processes allows to identify promising working fluid candidates together with optimal cycle conditions.

However, such computer simulations are subject to modelling uncertainties due to the operational conditions, process correlations and fluid properties. In this thesis the focus lies on the uncertainties from physical and chemical property data, caused by the experimental measurements or by the prediction models.

This thesis project presents a comprehensive framework to assess property uncertainties for different levels of thermodynamic cycle models. The framework consists of 1) a methodology for the development and uncertainty analysis of group contribution based property models, 2) a Bootstrap method for the quantification of uncertainties associated to equations of state parameters, 3) a Monte Carlo procedure for the propagation of property uncertainties through the cycle process onto the model output uncertainty, and 4) novel strategies for the selection of working fluids under property uncertainties, in particular a new reverse engineering approach based on sampling and uncertainty concepts. The framework is applied to different applications and case studies from industrial project partners.

Novel group contribution based property models are developed for the estimation of flammability-related properties (e.g. the lower flammability limit) of working fluids. Compared to existing models, the ones presented here show a higher accuracy, are simpler to apply and provide every prediction value with its corresponding uncertainty range (with 95% confidence). The study also reveals that

group contribution methods can suffer from parameter identifiability issues characterized by a significant correlation between estimated parameters. Hence, in order to ensure reliable estimation, reporting the 95% confidence interval of the model predictions is important.

In a second application it is shown how the uncertainty propagation of two types of equations of states, cubic and PC-SAFT, can be compared in the context of an industrial organic Rankine cycle, used for the recovery of waste heat from an engine of a marine container ship. The study illustrates that the model structure is vital for the uncertainties of equations of state and suggests that uncertainty becomes a criterion (along with e.g. goodness-of-fit or ease of use) for the selection of an equation of state for a specific application.

Furthermore, two studies on the identification of suitable working fluids for thermodynamic cycles are presented. The first one selects and assesses working fluid candidates for an organic Rankine cycle system to recover heat from a low-temperature heat source. The ranking of working fluids can be significantly different based whether the mean value or the uncertainties (e.g. the lower bound of the 95%-confidence interval) of the model output are considered. Hence, uncertainty analysis with respect to the input property uncertainties is a vital tool for model analysis and fluid selection.

In the second fluid selection study the novel reverse engineering approach based on sampling techniques and uncertainty analysis is applied to identify suitable working fluids for a industrial heat pump system, used to recover heat from spray-drying air in dairy industries. The novel reverse engineering approach provides a valid alternative to computationally demanding optimization approaches and allows to take into account property uncertainties.

The outcome of this thesis asserts that property uncertainties should be taken into account for process simulation applications, in order to support the model-based and reliable decisions on process fluids and process design.

Resumé på dansk

Der er en stigende industriel interesse for at genvinde spildvarme der udledes ved lave temperaturer (70-250°C). Termodynamiske kredsprocesser, såsom varmepumper og Rankine kredsprocesser, kan genvinde sådanne spildvarmer. Disse kan siden overføres til andre processtrømme eller konverteres til elektricitet.

Fluiderne, der cirkulerer i kredsprocesserne, er vigtige for kredsprocessens virkningsgrad. Modellering af fluidernes egenskaber samt kredsprocessen gør det muligt at identificere nye fluider samtidig med at procesbetingelserne optimeres. Computersimuleringers forudsigelser er dog belagt med usikkerheder på grund af usikkerhed omkring driftsbetingelser, proceskorrelationer og arbejdsmediernes egenskaber.

I denne afhandling ligger fokus på det bidrag til usikkerheden fra fysisk og kemiske data, der hidrører fra de eksperimentelle målinger der ligger til grund for de anvendte data, eller de modeller der anvendes til forudsigelse af data. Projektet skaber en ramme for vurdering af den indflydelse fysiske og kemiske egenskabers usikkerhed har på modeller for termodynamiske kredsprocesser. Denne ramme udgøres af 1) en metode til bestemmelse af prediktionsintervaller for gruppebidragsmetoder, 2) en Bootstrap metode til bestemmelse af prediktionsintervaller for tilstandsligninger, 3) en Monte Carlo metode til bestemmelse af modellernes output varians baseret på (input) variansen af fysiske og kemiske data 4) principper til udvælgelse arbejdsmedier under hensyntagen til usikkerheder i egenskaber, ikke mindste en ny 'reverse-engineering' strategi baseret på testning og variansestimation.

Metoderne testes i forbindelse med forskellige anvendelser og 'case'-studier med udgangspunkt i industrielle situationer.

Nye gruppebidragsmetoder udvikles til estimering af brandbarhedsrelaterede egenskaber (for eksempel den nedre brændbarhedsgrænse) for arbejdsvæsker. Sammenlignet med eksisterende modeller viser de nye metoder højere nøjagtighed.

De nye metoder er nemmere at anvende og hver forudsigelse med tilsvarende prediktionsinterval (95% konfidens) omkranser de eksperimentelle værdier. Undersøgelsen afslører også, at gruppebidragsmetoder kan have problemer med parametres identificerbarhed, kendetegnet ved en signifikant korrelation mellem estimerede parametre. Derfor er det vigtigt at rapportere det 95% konfidensinterval for at sikre en pålidelig belysning af disse forhold.

Et andet eksempel viser, hvorledes variansen af forudsigelser foretaget med to typer af tilstandsligninger, kubisk og PC-SAFT-ligningen kan sammenlignes. En industriel Rankine kredsproces, til genvinding af spildevarme fra en motor på et container skib, er udgangspunkt for analysen. Analysen viser, at modellernes parameteriseringsmetode er afgørende for variansen af tilstandsligningers estimater og antyder, at usikkerhed bliver et kriterium (sammen med for eksempel kurvetilpasning eller brugervenlighed) til udvælgelsen af tilstandsligninger til en bestemt anvendelse.

Desuden præsenteres to analyser til identifikation af egnede arbejdsmedier til termodynamiske kredsprocesser. Den første vælger og vurderer alternative arbejdsmedier til en Rankine kredsproces der skal genvinde varme fra en varmekilde ved lav temperatur. Arbejdsmedier rangeres væsentligt anderledes, hvis usikkerheder (for eksempel den nedre grænse for 95%-konfidensintervallet) på procesmodellens middelværdi betragtes. Derfor er analyse af prediktionsintervaller et vigtigt instrument i forbindelse med proces og arbejdsmedie-design.

I den anden undersøgelse med henblik på udvægelse af arbejdsmedie, anvendes den nye 'reverse engineering' tilgang baseret på testning og variansestimation. Dette gøres med henblik på identifikation af egnede arbejdsmedier til et industrielt varmepumpeanlæg, der bruges til at genvinde varme fra tørreluft i et spraytørringsanlæg. Den nye 'reverse engineering' tilgang udgør et alternativ til beregningsmæssigt krævende optimeringsmetoder og gør det muligt at tage hensyn til usikkerhed i fysiske og kemiske data.

Det er en konklusion af denne afhandling at usikkerheder af fysiske og kemiske data kan (og bør) tages i betragtning for at understøtte processimulering og andre modelbaserede beslutninger om arbejdsmedier og procesdesign.

Preface

This project was conducted in the scope of the THERMCYC Project, a collaborative research project between the Department of Chemical and Biochemical Engineering and the Department of Mechanical Engineering of the Technical University of Denmark (DTU), as well as several industrial partners, among others Viegand & Maagøe, MAN Diesel and Turbo, A. P. Møller-Mærsk and Arla Foods.

The thesis is organized in two parts:

- The first part summarizes the methodologies, case studies and findings of the thesis project.
- The second part consists of the articles published in peer-reviewed scientific journals and conference proceedings in the period 2014-2017.

There are many people who have contributed to this work, and it has been a great pleasure working with everyone involved over the past three years.

First and foremost, I would like to thank my supervisors, Gürkan Sin and Jens Abildskov. I deeply appreciate your guidance, your support, all our constructive discussions as well as all scholarly inputs throughout this project.

I also would like to express great thanks to all people involved in the THERMCYC Project, in particular to all my colleagues: To Jesper Graa Andreasen, for taking the initiative to our successful collaborations and specially for supporting me in modeling organic Rankine cycles and for explaining me all the details about the case studies. To Wei Liu, for his important help when building a model for cubic equations of states. To Stefano Cignitti, for all great discussions about new ideas for working fluid selection and design. To Benjamin Zühlsdorf for his valuable support during the heat pump case study. A special thank also to Fridolin Müller Holm for formulating the case studies together with the industrial partners and to Brian Elmegaard for leading and coordinating the THERMCYC project.

I am very thankful to Prof. John O'Connell for his valid contribution in the study on uncertainty propagation of equations of state.

In addition, I would like to thank Kenneth Kroenlein for welcoming me in his group at NIST for the external stay. Special thanks goes to Ian Bell for his valuable inputs to my work and for all great discussions.

I am also very grateful to Camille Marcarie for the project work during her stay at DTU, and to Merve Öner for helping me proofreading the thesis.

Thanks to all friends at DTU for an encouraging working atmosphere and for all fun events.

Last but not the least, I would especially like to thank my parents and Katja for their love, support and encouragement.

Jérôme Frutiger
Kgs. Lyngby, May 2017

Contents

Part I

Chapter 1

Introduction

This chapter contains a general overview of the background and the motivation of the PhD project as well as the structure of the thesis. Furthermore, the main contributions and the dissemination activities are briefly summarized.

1.1 Background

The efficient use of energy resources is important in all industrial applications, in particular to cost-effectively use electricity or fuel and to limit environmental impacts. Currently, there is still a large potential for the recovery of waste heat at low temperatures (70-250°C) [1]. In Danish industry such waste heat corresponds to up to 13% of the total energy consumption [2].

Low temperature waste heat can be found in all domains of industry. For example in

- evaporation, distillation and refrigeration processes of chemical plants [2]
- exhaust gas of large container ship vessels [3]
- drying processes in food industries [4]
- refrigeration and heating of buildings [5].

Thermodynamic cycles allow the recovery of waste heat. Two commonly used configurations of cycles are the heat engine and the heat pump (see Figure 1.1). In a heat engine the recovered heat is converted into electrical energy. An example of a heat engine is an organic Rankine cycle (ORC), which consists of a pump, an evaporator, a turbine and a condenser. The main components of a heat pump are a compressor, a condenser, an expansion valve and an evaporator. Heat pumps use electrical energy in order to transfer the recovered heat to a hot process stream. For both configurations a working fluid circulates and facilitates the heat transfer.

The potential of thermodynamic cycles for waste heat recovery can be understood by the following two examples: An ORC used to recover the exhaust gas heat of large container ship can improve the electrical output of the main engine power by up to 11% [6]. In this way the ship can safe fuel and decrease the amount of emitted CO_2 per kilometre of transportation. A heat pump for the recovery of heat from drying processes in food industry can decrease the utility costs by up to 36% [4].

Computational models are used for analyzing and optimizing the performance of thermodynamic cycles. In this context it is vital to consider the component design, the operating conditions and the influence of the working fluid [7].

This thesis project is situated in the domain of computational modelling of working fluid properties and selecting suitable working fluid candidates for case studies considering low-temperature heat recovery in industrial applications.

In the early design stage, database screening and molecular design techniques can be applied to test and evaluate promising pure component and mixture working

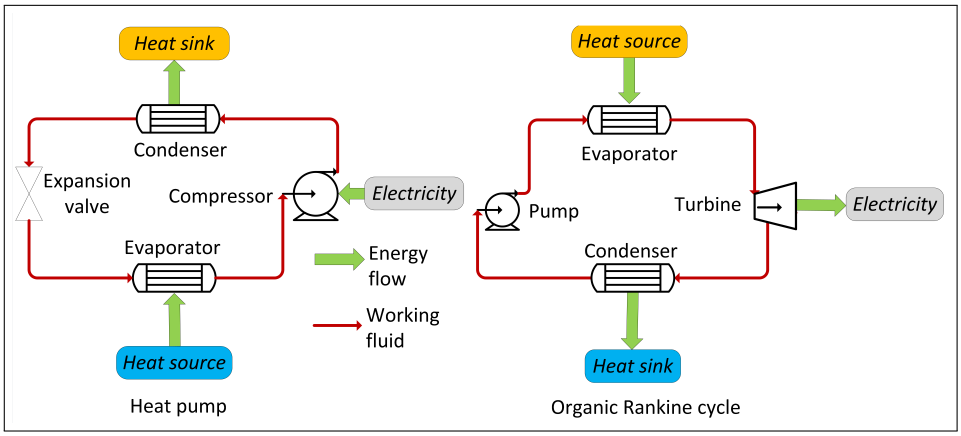


Figure 1.1: Basic configuration of a heat pump and an organic Rankine cycle (ORC).

fluids to optimize the cycle performance. The reviews on fluid selection provided by Bao et al. [8] and fluid design by Linke et al. [9] give a broad overview of the literature, which is available on techniques to identify working fluids for thermodynamic cycles. Recently the combination of fluid design and selection techniques with cycle process optimization has become established as a promising approach to achieve high cycle efficiencies [10].

The majority of working fluids and refrigerants used in thermodynamic cycles have been fluorinated and chlorinated compounds [11]. However, these fluids are subject to gradual phase-out due to high ozone depletion potential and global warming potential [12]. In recent years there is a particular focus on identifying working fluids with no ozone depletion and a low global warming potential. However, many fluids that satisfy these requirements (e.g. hydrocarbon based working fluids) are highly flammable. Hence, the adequate prediction and measurement of flammability-related properties become important [13].

Database searches or molecular design algorithms rely crucially on experimental and predicted property data. Property prediction models are especially used in the form of equations of state (EoS) to estimate thermodynamic properties (i.e. enthalpy, entropy and fugacity) [14]. Further, group contribution (GC) based models are widely used to estimate primary fluid properties (e.g. critical properties) in the context of molecular design studies.

However, property data and prediction models are subject to uncertainty, e.g. caused by the measurements [15] or by the property prediction models [16]. There are also other modelling uncertainties, e.g. caused by correlations in cycle component models or due to an incomplete knowledge of the operational conditions (e.g. varying heat source temperatures) [17, 18]. The current thesis focuses though on

property uncertainties of pure components.

There is a difference between accuracy and uncertainty in the context of computational models for property prediction and process design. Accuracy is the difference between the output predicted by the model and a particular set of experimental measurements of a property or process output [19]. Uncertainty is the range of statistically possible outcomes of the model (usually assumed to be a normal distribution and reported with 95% confidence) [20]. This difference is visualized in Figure 1.2 for property data.

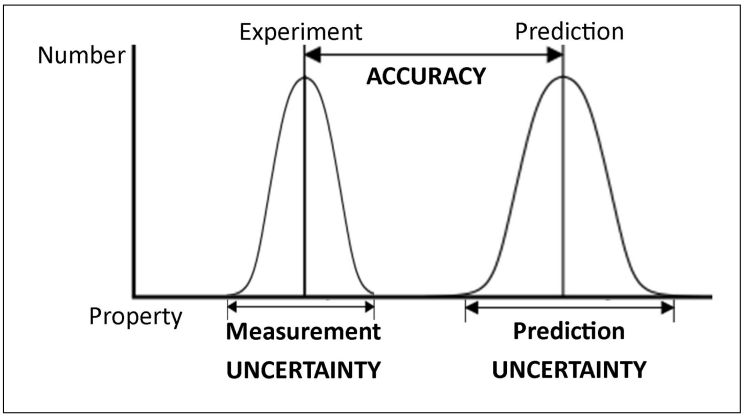


Figure 1.2: Illustration of uncertainty and accuracy in property modelling. Accuracy is the difference between the output predicted by the model and the experimental measurement of the property or process output. Uncertainty is the range of statistically possible outcomes of the property prediction model. Also shown in the picture is the measurement uncertainty, centered around the experimental (mean) value.

In the scope of good modelling practices, it is necessary to take these property uncertainties into account in order to establish the application range and the reliability of the overall design model for thermodynamic cycles [20].

Figure 1.3 gives an overview of the inter-connection between property models and a thermodynamic cycle model and also illustrates, how property uncertainties can propagate through the cycle.

The basis of all property prediction models are experimental measurements. As an example experimental vapour pressure data can be used to fit parameters of an EoS or the parameters of property prediction models, such as GC methods. Both experimental values and property prediction methods can be used to obtain primary fluid properties, such as critical properties, which can serve as input properties to an EoS (e.g. in the case of cubic EoS [14]). Furthermore, safety-related and environmental properties can be predicted [21] whenever the corresponding experimental value is not available, e.g. due to the novelty of a certain fluid.

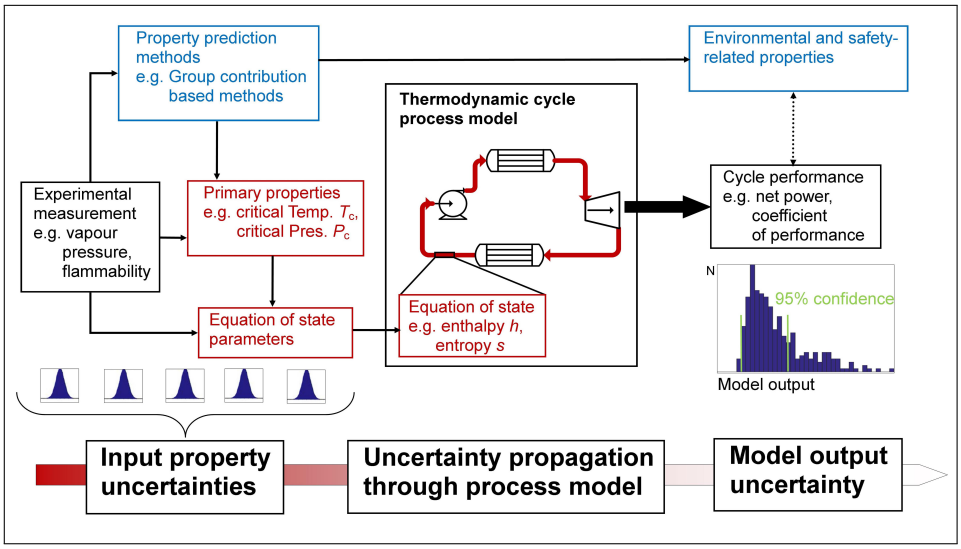


Figure 1.3: Illustration of the connections between property parameters, property models and process models, as well as overview of input property uncertainties and their propagation through the thermodynamic cycle model.

For property models the sources of uncertainties [20] can be generally found in:

- the property model parameters obtained through fitting to experimental data,
- the mathematical formulation of the property model that only approximates nature, and
- the stochastic components of a process simulation.

As shown by the scheme of Figure 1.3, the property models are connected to one another and to the process model. Hence, the input property uncertainties can propagate through the EoS into the cycle model and onto the cycle model output (e.g. power for an ORC or the coefficient of performance for a heat pump).

A thorough uncertainty analysis can give an additional dimension during decision process for the design of a thermodynamic cycle and the selection of suitable working fluids: e.g. a conservative or an optimistic decision approach, corresponding to the lower or upper bound of the uncertainty range (i.e. the confidence interval) [20] .

It is also possible to analyze the applied property models used for the cycle, e.g. to assess a chosen EoS and compare it to alternatives based on the property uncertainty analysis [22].

The quantification of the influence of property parameter uncertainties on the model output, i.e. through a sensitivity analysis [23], can also contribute to the

understanding of the cycle process or support the identification of major fluid target properties.

However, there is still a lack of application of property uncertainty analysis tools in complex design problems, as in the domain of working fluid selection and design [24]. So far only a few studies have been conducted that addressed property uncertainties in thermodynamic cycle models: Papadopoulos et al. [25] included a nonlinear sensitivity analysis method to address model-related uncertainties in the mixture selection procedure. The sensitivity analysis method of Papadopoulos et al. was specifically adapted by Mavrou et al. [26] for the identification of optimal fluid mixtures under changing design and operating parameters.

In the case of environmental and safety-related properties, there have been efforts to predict and quantify the global warming potential and the ozone depletion potential [21], but there is no prediction method for flammability-related properties of working fluids that specifically includes an uncertainty statement [27, 28].

1.2 Motivation and goals

The aim of this thesis project is to provide a comprehensive framework with methodologies and tools to assess property uncertainties on different levels of a thermodynamic cycle model (as visualized in Figure 1.3). Furthermore, the project should advance the field, by showing 1) how these tools can be utilized for industrial applications, and 2) what particular insights can be gained when property uncertainties are taken into account.

Specifically the following major challenges motivated the current thesis project:

- Working fluid design or selection studies usually neglect the impact of property uncertainties on the cycle model output [25] and there are no established methodologies to take property uncertainties into account, while identifying suitable working fluids.
- Uncertainty analysis is generally not used as a criterion to assess, compare or select EoS for a certain application [22].
- GC based property methods suffer from major numerical and statistical issues [16], e.g. outliers and parameter correlation, which influence the prediction and its corresponding uncertainty. These problems have not yet been addressed in a systematic way and discussed sufficiently, although the work of Hukkerikar et al. [16] represents a beginning.

In order to systematically address these challenges, the specific goals of this research project are the following:

1. A methodology to estimate the parameters and quantify the uncertainties of GC based property models should be established and utilized for the development of safety-related properties of working fluids. The statistical issues related to the uncertainty of GC based parameters should be investigated and suggestions should be made for developers and users of GC based methods.

2. A methodology to quantify the uncertainties of parameters of established EoS from experimental data should be described. Furthermore, a procedure should be suggested to propagate property uncertainties onto the cycle process model output. These computational tools should allow the systematic assessment and comparison of the uncertainty propagation of EoS in the context of an industrial thermodynamic cycle.

3. Novel strategies to identify working fluids for thermodynamic cycle applications under property uncertainties should be suggested. These strategies should be demonstrated through case studies concerning low-temperature heat recovery in the Danish industry and promising working fluids should be suggested.

The demonstration of the framework through applications in the context of working fluids for thermodynamic cycles should also encourage users and developers of other process simulation applications to analyse and state the process output uncertainty due to chemical property uncertainties.

1.3 Outline of the thesis

The thesis is structured in two parts. The first part is the main body text, which introduces the principles and contains the main research results. The second part is the collection of publications (journal articles and peer-reviewed conference proceedings), which contains all the details about the methodologies and case studies.

The first part starts in Chapter 2 with the description of a framework containing computational methodologies and tools to analyze and propagate property uncertainties.

The methodologies of the framework have been used for applications and case studies, that are outlined in Chapter 3. There are two applications concerning the uncertainty analysis of property models (i.e. GC methods and EoS). With respect to the analysis of EoS, the methodology is applied in the context of an ORC to recover heat from the exhaust gas of a marine transportation vessel. Furthermore, two case studies concern the selection of working fluids for a thermodynamic cycle system. The first case study is an ORC for the recovery of heat from a hot water process

stream. The second case study considers a heat pump system for the recovery of heat from hot air used in spray drying facilities of a milk powder production plant.

Chapter 4 shows the results of the uncertainty analysis of the property models and Chapter 5 describes the outcomes of the fluid selection studies. Finally, Chapter 6 puts the results in perspective, states the major conclusions of the thesis and gives suggestions for future research studies.

1.4 Summary of main contributions

The thesis is based on methodologies, case studies and findings of that have been published in journal articles and peer reviewed conference proceedings, which are all included in Part II as **Papers A** to **H**.

(**A**) **Journal article**:
Jérôme Frutiger, Camille Marcarie, Jens Abildskov, Gürkan Sin: "A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models - An application to the heat of combustion", *Journal of Chemical and Engineering Data*, 61, 1, pages: 602-613, 2016.

This work describes in detail a rigorous methodology that addresses numerical and statistical issues while developing GC based property models such as regression methods, optimization algorithms, performance statistics, parameter identifiability and uncertainty of the prediction. The statistical analysis of the GC method reveals new insights in parameter identifiability issues associated with GC methods and the inclusion of additional parameters (e.g. higher order groups) and suggests that the 95%-confidence intervals of the predicted property values should be mandatory as opposed to reporting only single value predictions.

(**B**) **Peer reviewed conference proceedings**:
Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "Outlier treatment for improving parameter estimation of group contribution based models for upper flammability limit", *Proceedings of the 25th European Symposium on Computer Aided Process Engineering ESCAPE 25*, 37, pages: 503-508, 2015.

The study provides an additional insight to one aspect of the methodology described in (**A**): A systematic method for outlier treatment in order to im-

prove the parameter estimation of GC models. The new method identifies and removes outliers based on the empirical cumulative distribution function.

(**C**) **Journal article**:
Jérôme Frutiger, Camille Marcarie, Jens Abildskov, Gürkan Sin: "Group-contribution based property estimation and uncertainty analysis for flammability-related properties", *Journal of Hazardous Materials*, 318, pages: 783–793, 2016.

The methodology described in paper **A** and **B** has been applied to develop new GC models for the prediction of safety-related properties of organic process chemicals and working fluids (e.g. the lower and upper flammability limits. Every estimated property value is reported together with its confidence of the prediction. Compared to existing models, the developed ones have a higher accuracy, are simple to apply and provide uncertainty information on the calculated prediction (i.e. the 95%-confidence interval).

(**D**) **Journal article**:
Jérôme Frutiger, Ian Bell, John P. O'Connell, Kenneth Kroenlein, Jens Abildskov, Gürkan Sin: "Uncertainty assessment of equations of state with application to an organic Rankine cycle", *Molecular Physics*, 115, 1, pages: 1-20, 2017.

The article presents the first generic methodology to analyse and compare equations of state (EoS) from a detailed uncertainty analysis of the mathematical form and the data used to obtain EoS parameter values. The procedure is illustrated by comparison of cubic Soave–Redlich– Kwong (SRK) EoS with perturbed-chain statistical associating fluid theory (PC-SAFT) EoS for an ORC using cyclopentane as working fluid. Uncertainties of the EoS input parameters including their corresponding correlation structure are quantified from experimental measurements using a bootstrap method. A Monte Carlo procedure propagates parameter input uncertainties onto the ORC model output.

(**E**) **Peer reviewed conference proceedings**:
Jérôme Frutiger, Ian Bell, John P. O'Connell, Kenneth Kroenlein, Jens Abildskov, Gürkan Sin: "Uncertainty assessment of equations of state with application to an organic Rankine cycle", *Proceedings of the 30th International Conference on Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems ECOS-2017*, 2017.

This paper is based on the previous work presented in paper (**D**) for the

quantification of the uncertainties of EoS parameters and their propagation onto the ORC model output. The study thoroughly compares cubic SRK, PC-SAFT and the cubic Peng-Robinson EoS. The EoS were analysed using two parametrizations, the one common in the literature (critical properties and acentric factor) and one re-parametrized form that allows fitting to the same experimental data as PC-SAFT. The common parametrizations of the SRK and PR EoS models yielded narrower uncertainty distributions than PC-SAFT. However, when fitted to the same data as PC-SAFT, the uncertainty distributions for the model output became much broader for SRK and PR, in fact close to the values of PC-SAFT. This suggests that extensive re-fitting of parameters to data, will amplify the uncertainty in these data.

(**F**) **Journal article**:
Jérôme Frutiger, Jesper Andreasen, Wei Liu, Hartmut Spliethoff, Fredrik Haglind, Jens Abildskov, Gürkan Sin: "Working fluid selection - Impact of uncertainty of fluid properties", *Energy*, 109, pages: 987-997, 2016.

This article presents a generic methodology to select working fluids for an ORC system taking into account property uncertainties of the working fluids. The methodology has been applied to a working fluid selection problem for an ORC using a low-temperature heat source. After screening of 1965 possible working fluid candidates, the uncertainties of the fluid properties were specified and propagated using a Monte Carlo procedure to the ORC model output, providing the 95%-confidence interval of the net power output of the cycle for every fluid. The methodology proposes fluid property uncertainties as an additional dimension to the fluid selection process.

(**G**) **Journal article**:
Jérôme Frutiger, Benjamin Zühlsdorf, Brian Elmegaard, Jens Abildskov, Gürkan Sin: "Reverse engineering of working fluid selection for industrial heat pump based on Monte Carlo sampling and uncertainty analysis", *Energy*, in preparation.

This article presents a new strategy to solve product-process design problems using Monte Carlo sampling based reverse engineering. Monte Carlo sampling technique is used to sample property values creating a set of virtual fluid candidates, which is evaluated in the process model. Ranking of the solutions provides the optimal target property values to look for in a database search. The real fluids closest to the best virtual fluids are identified by searching in

chemical databases, and evaluating a distance function between real and virtual fluid property values. The closeness of a real fluid to a virtual fluid is calculated taking into account the uncertainty range of the real fluids properties. The method has been demonstrated for the identification of the best suitable working fluid in a heat pump system used to recover low-temperature heat in spray-drying facilities of food industries.

(H) Peer reviewed conference proceedings:
Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "Global sensitivity analysis of computer-aided molecular design problem for the development of novel working fluids for power cycles", *Proceedings of the 26th European Symposium on Computer Aided Process Engineering ESCAPE 26*, 38, pages: 283–288, 2016.

This study compares two methods for global sensitivity analysis as a new approach for the identification and ranking of target properties in molecular design problems: A modified Morris Screening technique and Monte Carlo based standard regression. The two methodologies are highlighted in a case study involving the design of a working fluid for an ORC design for power generation.

Conference presentations

Contributions were also disseminated through specialized session talks in the following conferences:

- Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "Estimation and uncertainty analysis of flammability properties for computer-aided molecular design of working fluids for thermodynamic cycles", CAPE Forum, 2015, Paderborn.
- Jérôme Frutiger, Jens Abildskov, Gürkan Sin: " Estimation and uncertainty Analysis of flammability properties of chemicals using group-contribution property models", 12th PSE and 25th ESCAPE Joint Conference, 2015, Copenhagen.
- Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "A rigorous methodology for development and uncertainty analysis of group contribution based property models", Nineteenth Symposium on Thermophysical Properties, 2015, Boulder.
- Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "Sensitivity analysis of computer-aided molecular design problem for the development of novel working fluids

for power cycles", 10th European Congress of Chemical Engineering, 2015, Nice.

- Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "Sensitivity analysis of molecular design problem for the development of novel working fluids for power cycles", AIChE Annual Meeting, 2015, Salt Lake City.
- Jérôme Frutiger, Jens Abildskov, Gürkan Sin: "Group contribution modelling for the prediction of safety-related and environmental properties", AIChE 2016 Spring Meeting and 12th Global Congress on Process Safety, 2016, Houston.
- Jérôme Frutiger, John P. O'Connell, Jens Abildskov, Gürkan Sin: "Equation of state selection for organic Rankine cycle modeling under uncertainty", AIChE Annual Meeting, 2016, San Francisco.

Chapter 2

Framework for property uncertainty analysis and identification of optimal working fluids

This chapter describes the overall computational framework that has been developed in this research project. The framework assesses the property uncertainties on different levels. It starts with a methodology for parameter estimation and uncertainty analysis of group contribution based property models. Further, it is explained, how the uncertainties of equations of state parameters can be quantified from experimental data. Afterwards, it is illustrated how the quantified property uncertainties can be propagated through a cycle process model. Finally, novel strategies are suggested for fluid selection as well as for a reverse engineering approach, based on sampling and uncertainties.

2.1 Development, parameter estimation, and uncertainty analysis of group contribution based property models

Group contribution (GC) methods use a function of structurally dependent parameters. GC methods can be used for the prediction of working fluid properties, whenever experimental values are missing. The prediction and the corresponding uncertainty range can be obtained for primary properties [16], but also for safetety-related [29] and environmental properties [21].

Here it is shown, how the property uncertainties for GC based property prediction methods can be quantified and assessed. A comprehensive and step-by-step methodology for the development of GC based property models is described here. The detailed description of the methodology can be found in **Paper A** in Part II.

An overview of the methodology including the workflow, the data and techniques used at each step is shown in Figure 2.1. The methodology has been developed based on the work of Hukkerikar et al. [16].

2.1.1 Property model structure definition and choice of regression method

Here the Marrero/Gani (MG) [30] method is selected. This method combines the contributions from a specific functional group (1st order parameters), from polyfunctional (2nd order parameters) as well as from structural groups (3rd order parameters). By using higher order parameters (2nd and 3rd), additional structural information about molecular fragments is provided. The general form of the MG method is

$$f_i(X) = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l \tag{2.1}$$

$$f(X) = T \cdot \theta \tag{2.2}$$

In Eq. (2.1) C_j is the contribution of the 1st order group of type j that occurs N_j times whereas D_k is the contribution of the 2nd order group of type k that occurs M_k times in the molecular structure of a pure component. E_l is the contribution of the 3rd order group of type l that has O_l occurrences. The function $f(X)$ is specific for a certain property X. The parameters can be collected in the vector θ and the occurrences of the groups can be depicted in the matrix T as shown in Eq. (2.2). As an example, the different GC-factors of 1,2-Dichloro-4-nitrobenzene and Adiponitrile are visualized in Figure 2.2.

A high number of experimental data points is a prerequisite in order to obtain an accurate model with a wide application range. Experimental property data can

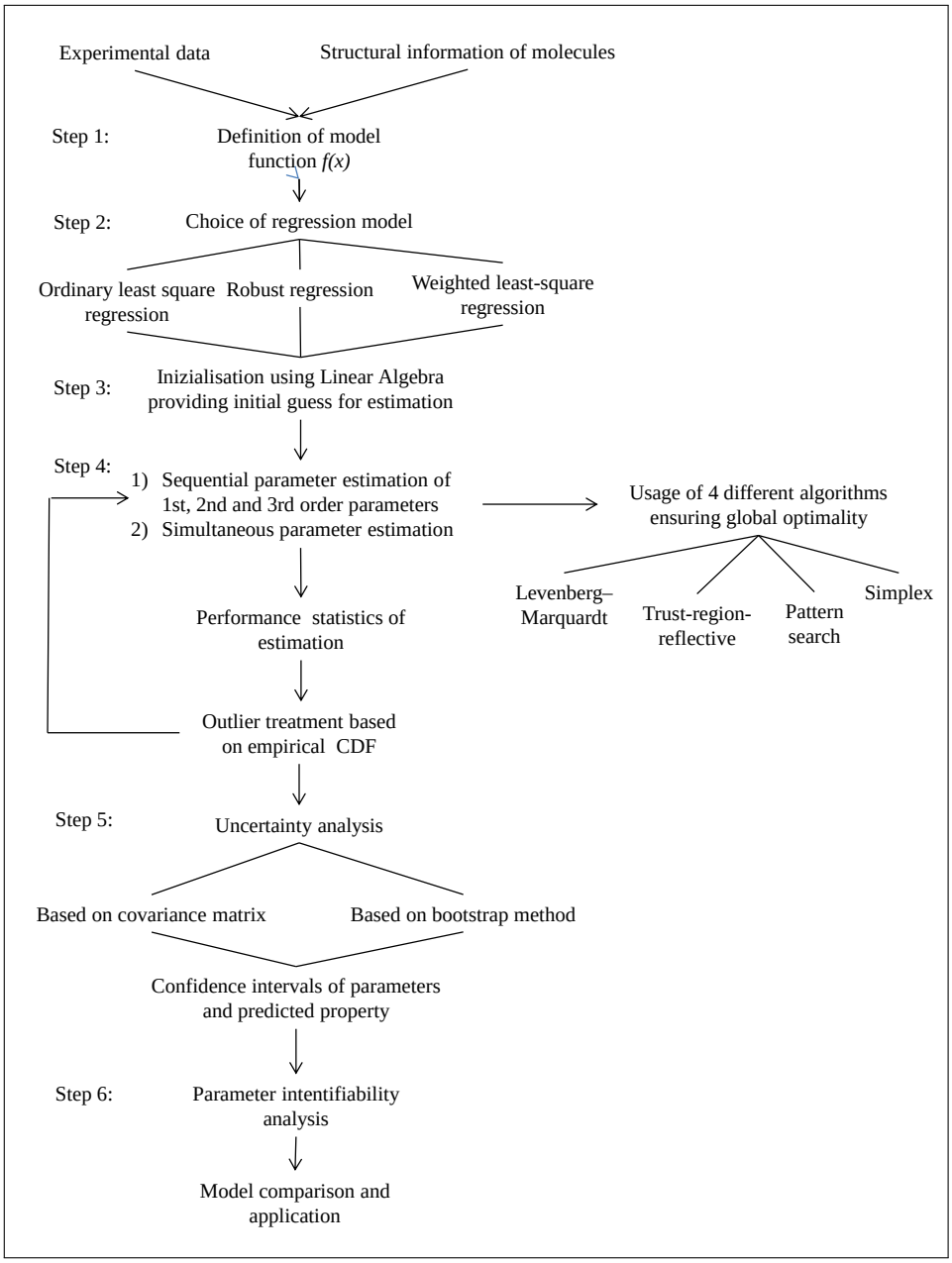


Figure 2.1: Overview of the methodology for development, parameter estimation and uncertainty analysis of GC based property models.

be found in data bases such as AIChE DIPPR 801 Database [31], NIST ThermoData Engine (TDE) [32, 33] or DECHEMA DETHERM [34] or CHEMSAFE [35].

After assigning the different 1st, 2nd and 3rd order groups to the respective

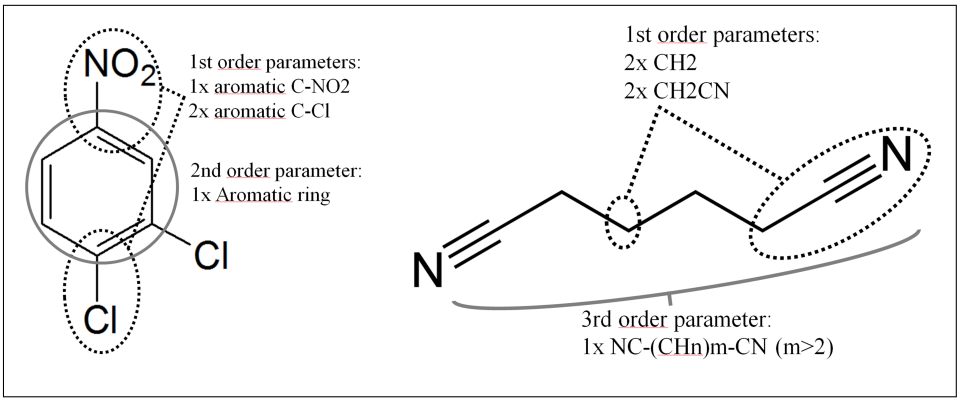


Figure 2.2: Example of Marrero/Gani group contribution factors of 1,2-Dichloro-4-nitrobenzene (left) and Adiponitrile (right).

molecules, it is necessary to determine a model function. We seek a function of the property which is linear in the contribution factors. Hence, a suggestion for the property function is obtained by generating plots of various classes of pure components versus their increasing carbon number in homologue series as already shown by Pierotti et al. [36]. As an example a selection of classes of compounds is shown for the heat of combustion ΔH_c° (see Figure 2.3). From these plots, a linear function is deemed as appropriate model function for the ΔH_c° property.

Three regression models have been investigated in this work for the use in parameter estimation in GC model development.

- Robust regression [37]
- Ordinary nonlinear least squares regression [38]
- Weighted nonlinear least squares regression [39]

In robust regression each residual is weighted by a certain factor w_i [38], placing high weights on small residuals and small weights on large residuals. In this way the influence of data points producing large residuals (not following the model), i.e. potential outliers, is decreased [37].

In ordinary nonlinear least squares regression the weights are one for all data points ($w_i = 1$) and in weighted nonlinear least squares regression the weights are equal to the inverse of the corresponding variances of the experimental measurement of the corresponding data point [39].

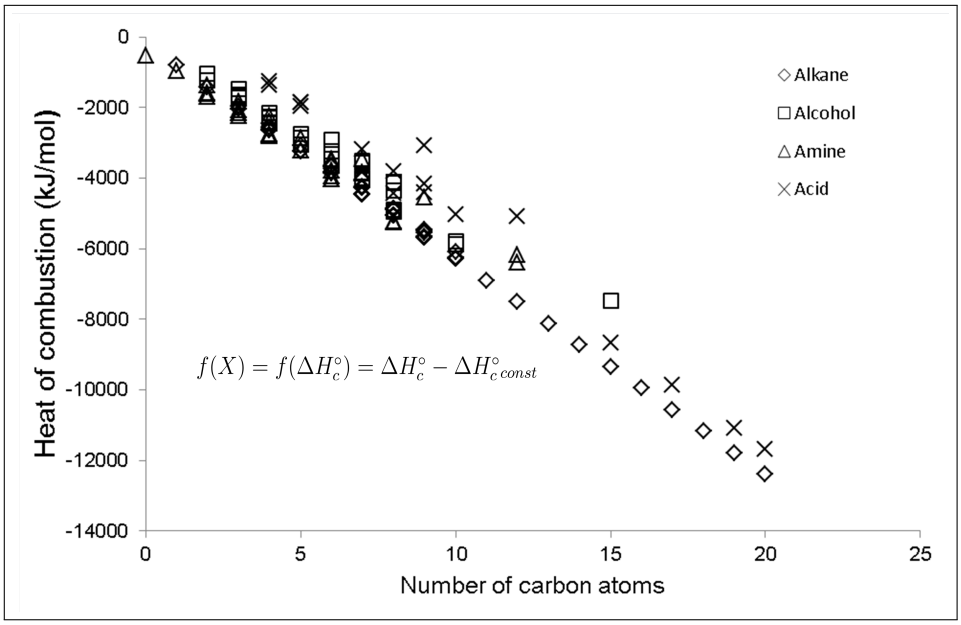


Figure 2.3: Graphical analysis of number of carbon atoms versus property to infer about a proper model function: (y-axis) heat of combustion ΔH_c° of a selection of pure components, (x-axis) carbon number of pure components in increasing order.

2.1.2 Sequential and simultaneous parameter estimation and outlier treatment

The 1st, 2nd and 3rd order parameters are first estimated separately (i.e. sequentially) applying the non-linear regression model chosen before. The result of the sequential estimation serves as initial guess for the simultaneous parameter estimation algorithm, where all parameters are estimated together for the chosen regression problem. In order to test that the global minimum of the least-squares regression is achieved, a practical approach is followed, in which 4 different optimization algorithms are applied: Levenberg–Marquardt algorithm [40], trust-region reflective algorithm [41], simplex algorithm [42], pattern search optimization [43].

The performance of the parameter estimates should be quantified by a variety of statistics in order to obtain a broad set of measures [16], for example the Pearson correlation coefficient R^2 or the average relative error *ARE* between the predictive and the experimental value. The definition of statistical performance indicators can be found in **Paper (A)** in Part II.

The GC parameter estimation can be strongly influenced by outliers from the model structure. Here outliers are detected based on the empirical cumulative distribution function (CDF) of the residuals between experimental and predicted

values. This methodology is described in detail in **Paper (B)** in Part II for the identification of outliers in GC models.

The empirical CDF is a step function that increases by $1/N_{data}$ in every data point, where N_{data} is the number of data points. In this way, it seeks to estimate the true underlying distribution function of residuals and thereby improve the detection of outliers. It does not assume that residuals follow a normal distribution. Data points that lie below the 2.5% or above the 97.5% probability levels which corresponds to two standard deviations in normal distribution, are taken to be outliers. Figure 2.4 shows an example of the empirical CDF of the GC parameter estimation for the heat of combustion.

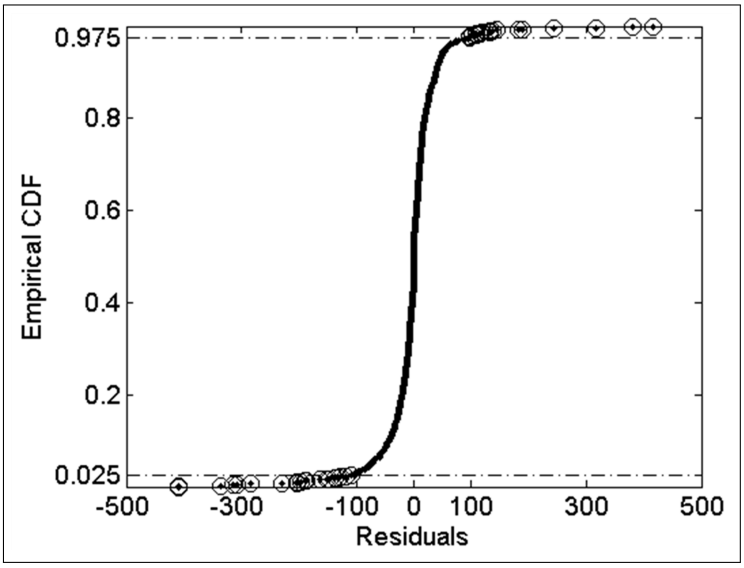


Figure 2.4: Empirical CDF of the residuals obtained from the GC parameter estimation for the heat of combustion. Below a probability of 0.025 and above 0.975 the data points are considered to be outliers.

2.1.3 Uncertainty analysis based on linear error propagation using parameter covariance matrix

The underlying assumption of this method for uncertainty analysis is that the measurement errors are ideally and independently distributed and defined by a Gaussian distribution (normal distribution with zero mean and unit standard deviation) [38]. The uncertainty of the parameter estimates is based on the asymptotic ap-

proximation of the covariance matrix, $COV(\theta^*)$, of parameter estimators [44]

$$COV(\theta^*) = \frac{SSE}{N_{data} - p}(J(\theta^*)^T J(\theta^*))^{-1} \tag{2.3}$$

In Eq. (2.3) SSE is the minimum sum of squared errors obtained from the least-squares parameter estimation method, N_{data} is the number of data points and p the number of parameters. The Jacobian J is the local sensitivity of the property model f with respect to the parameter values θ^*. The corresponding elements of the parameter correlation matrix can be obtained by

$$Corr(\theta_i^*, \theta_j^*) = \frac{COV(\theta_i^*, \theta_j^*)}{Var(\theta_i^*)Var(\theta_j^*)} \tag{2.4}$$

In Eq. 2.4 $COV(\theta_i^*, \theta_j^*)$ is the respective element of θ_i^* and θ_j^* of the covariance matrix and $Var(\theta_i^*)$ and $Var(\theta_j^*)$ are the variances of the respective parameters. The errors on property predictions are estimated using linear error propagation, in which the covariance matrix of the predictions, $COV(y^{pred})$, is approximated using the Jacobian and the covariance of the parameter estimates as shown in Eq. (2.5).

$$COV(y^{\text{pred}}) = J(\theta^*)COV(\theta^*)J(\theta^*)^T \tag{2.5}$$

A student t-distribution $t(n-p, \alpha_t/2)$ (with $\alpha_t/2$ percentile) can be used to calculate the confidence intervals of the parameters and the property predictions:

$$\theta_{1-\alpha}^* = \theta \pm \sqrt{\text{diag}(COV(\theta^*))} \cdot t(n-p, \alpha_t/2) \tag{2.6}$$

$$y_{1-\alpha}^{\text{pred}} = y^{\text{pred}} \pm \sqrt{\text{diag}(COV(y^{\text{pred}})} \cdot t(n-p, \alpha_t/2) \tag{2.7}$$

In Eq. (2.6) and (2.7) $\text{diag}(COV(\theta^*))$ and $\text{diag}(COV(y^{\text{pred}}))$ represents the diagonal elements of the respective matrices.

The assumption to follow normal distribution [38] is rarely the case in practice (see e.g. the residual plots in Hukkerikar et al. [16]). The bootstrap method is an attempt to calculate the distributions of the errors from the data, and to use these to calculate the errors on the parameter estimation [45].

The bootstrap method is described in detail in section 2.2 and is used there to quantify the uncertainties of parameters of equations of state. However, the method can easily be adapted for GC methods. As it is shown in **Paper (A)** in Part II, the bootstrap method can be considered as a valid alternative to classical uncertainty analysis.

Parameter identifiability is a common problem in nonlinear regression [38] with important implications for model validation and application. Parameter identifiability is basically the issue, whether the model parameters can be estimated uniquely

from a certain data set. We use the following diagnostic measures to analyze parameter identifiability in GC models:

(a) The parameter estimates *must not be linearly dependent*, so the linear correlation coefficients between parameter estimates should be sufficiently low, e.g. less than 0.7 [46, 47] and

(b) Parameter estimation errors (i.e. 95% confidence intervals) should be sufficiently low [48]. One obvious indication of poor parameter identifiability is a large confidence interval, e.g. relative parameter estimation error being larger than 50% [49].

2.2 Quantification of uncertainties of equation of state

A methodology that assesses the parameter uncertainty for any given equation of state (EoS) based on the experimental data is presented in the following. It consists of the quantification of uncertainty and the correlation structure of the EoS input properties and parameters. A detailed explanation of the procedure is given in **Paper (D)** and **Paper (E)** in Part II.

EoS are necessary to predict thermodynamic properties of working fluids such as enthalpies, entropies or fugacities at each state of a cycle. Experimental property data (e.g. critical properties, saturation pressures, and liquid densities) are normally used to determine parameters of an EoS. These data have associated uncertainties arising from the measurements [15] and how the model incorporates the values [16].

2.2.1 Formulation of equations of state

The thermodynamic properties enthalpy and entropy are expressed through an ideal contribution (i.e. the ideal-gas enthalpy and entropy) and a nonideal gas contribution (departure function) for the difference between ideal- and real-fluid behaviors [14]. In Eq. (2.8) the enthalpy expression is shown as an example:

$$h(P,T) = h^{\mathrm{ref}} + h(T)^{\mathrm{ideal}} + h(P,T)^{\mathrm{dep}} \tag{2.8}$$

where P is the pressure and T the temperature. The reference enthalpy, h^{ref}, is defined at the reference state (T^{ref}, P^{ref}). The ideal gas enthalpy is $h(T)^{\mathrm{ideal}}$, while $h(P,T)^{\mathrm{dep}}$ is the respective departure function. Fugacities can be directly calculated from EoS departure functions.

The ideal-gas enthalpy and entropy terms are usually obtained by integrating a temperature-dependent ideal-gas heat capacity function, $c_p(T)$. As an example the heat capacity can be written in a polynomial form ($c_p(T) = A + B \cdot T + C \cdot T^2 + D \cdot T^3$). The parameters of the ideal gas heat capacity are in that case: A, B, C, D.

There are a variety of different types of departure functions used for the description of working fluids [14, 50]: Forms of the Helmholtz EoS [51, 52], cubic EoS such as Peng-Robinson[53] or Soave-Redlich-Kwong (SRK) [54], or Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) [55].

As an example the non-associating PC-SAFT EoS can be considered, which has three parameters (the segment diameter σ, the energy parameter ε/k and the chain length parameter m) that must be obtained by fitting the EoS to a combination of property data, e.g. vapor pressure and (liquid) density data as function of temperature [55].

The goal of this methodology is to obtain the uncertainties and the correlation matrix of the respective EoS parameters for a certain chemical compound, e.g. for the PC-SAFT parameters (σ, ε/k, m). The quantification is based on the thermodynamic property data, which need to be collected from databases [32].

2.2.2 Bootstrap method for uncertainty analysis

The bootstrap method attempts to quantify the underlying distributions of residual errors commonly defined in statistical contexts as the differences between the experimental data and their corresponding model calculations. The residual errors are used to obtain synthetic data sets for parameter estimation by using random sampling with replacement. This procedure is a form of nonlinear propagation of measurement errors to errors in parameter estimators. It is different from nonlinear regression theory which relies on asymptotic approximation of the parameter covariance matrix that requires calculation of the jacobian matrix and the assumption that measurement errors are independently identically distributed and follow normal distribution with means equal to zero [38].

An outline of the bootstrap method [45] is provided here. A generic model $F[\theta]$ with parameters $[\theta]$ to predict variable y^{pred} is given by

$$y^{\text{pred}} = F([\theta]) \tag{2.9}$$

The goal is to fit the model parameters giving y^{pred} to the experimental data set, y^{exp} , of N_{data} data points, obtaining the parameter estimates θ^* and their corresponding uncertainties.

1. A reference parameter estimation is made using a non-linear least squares method to obtain the first parameter estimates θ^*:

$$\theta^* = \arg\min \sum_i (y_i^{\text{exp}} - F_i(\theta))^2 \tag{2.10}$$

2. The residual error for each data point is defined as:

$$\hat{\varepsilon}_i = (y_i^{\text{exp}} - y_i^{\text{pred}}) \tag{2.11}$$

   Each residual error has equal probability of occurring, with a probability of $1/N_{\text{data}}$.

3. New synthetic data sets are produced via the bootstrap method. Random sample replacements are made of residual errors $\hat{\varepsilon}$ to generate k synthetic data

sets ($y^*(1)$; $y(2)$, ..., $y(k)$), each with N_{data} data points. In practice, this bootstrap method simply samples errors and adds them randomly to the estimated properties in the reference step above (i.e., it rearranges the errors):

$$y_i(k)^* = y_i^{\text{pred}} + \hat{\varepsilon}_i \quad \hat{\varepsilon}_i \in \hat{F}(\varepsilon) \tag{2.12}$$

where i (from 1 to N_{data}) stands for the index of measured data and $\hat{F}$ is the probability function of $\hat{\varepsilon}$ (with probability of realization of $1/N_{\text{data}}$ for all $\hat{\varepsilon}$).

4. The least squares parameter estimation is repeated using each synthetic data set $y(k)$, which results in a new set of estimated parameters $\theta^*(k)$ and a new set of predicted values $y^{\text{pred}}(k)$. In this way, distributions of the parameters as well as of the predicted values are obtained for representing the uncertainty in the estimated values.

5. The obtained distributions of parameters represent their uncertainty and can be analyzed with interference statistics (such as the standard deviation) to quantify the uncertainty range.

As an example the bootstrap method is applied for the PC-SAFT parameters of cyclopentane, resulting in distributions for the σ, ε/k and m (see Figure 2.5). Collected experimental data for vapor pressure [56] and saturated liquid densities [57] have been used. The uncertainties are calculated from the ratio between calculated two standard deviations (SD) of the distributions and the actual value from the literature (see Table 2.1).

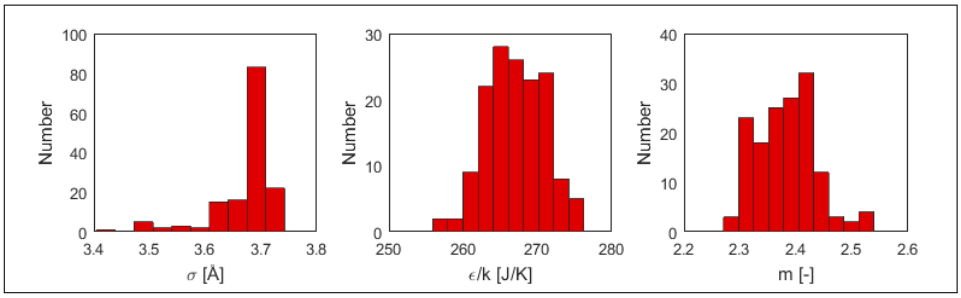


Figure 2.5: Distribution of PC-SAFT parameters obtained from bootstrap method.

The quantified uncertainties of the fluid-specific EoS parameters (see Table 2.1) can be used for the propagation through a thermodynamic cycle model in order to obtain the process model output uncertainty (see 2.3).

An important feature of the bootstrap method is that it allows estimation of the correlation structure between the errors of the different parameters (e.g. for PC-

SAFT: the correlation structure between the residual errors associated with values of σ, ε/k, and m is depicted in Table 2.2). It is essential for an uncertainty propagation to preserve the original correlation structure, in order to avoid the output uncertainty calculation being incorrect [58]. and In Table 2.2 ε/k and m are highly correlated, but σ is not correlated with the other two.

Table 2.1: Estimated uncertainties for PC-SAFT parameters in %, calculated from two standard deviations (SD) of distributions.

$\frac{2 \cdot SD(\sigma)}{\sigma}$	$\frac{2 \cdot SD(\varepsilon/k)}{\varepsilon/k}$	$\frac{2 \cdot SD(m)}{m}$
3.05%	2.89%	4.61%

Table 2.2: Correlation matrix of errors.

	σ	ε/k	m
σ	1		
ε/k	0.05	1	
m	-0.36	-0.94	1

Furthermore, the influence of different property uncertainty sources on the thermodynamic cycle outputs may be analyzed by comparing the different variances and standard deviations. As a result, it is possible to compare the effects of ideal-gas contributions uncertainties with those from the nonideal departure functions, or different types of EoS can be considered relative to eachother. To this extent sensitivity measures can be used as descibed by Saltelli et al. [23]. Detailed definitions and examples of sensitivity measures in the context of EoS for thermodynamic cycles can be found in **Paper (D)**.

2.3 Propagation of fluid property uncertainties to process model output uncertainty

In the following a Monte Carlo based procedure is described to propagate the influence of the input uncertainty of the fluid parameters on the model output of a thermodynamic cycle. Detailed explanations and an application example of the procedure can be found in **Paper (F)**.

The Monte Carlo based uncertainty propagation of fluid properties can be useful, when uncertainties of different property models are compared that are used in the cycle, e.g. to assess the uncertainty of different types of EoS on the cycle level (see **Paper (D)**). Furthermore, the propagated property uncertainties can give an additional dimension to the fluid selection process (see **Paper (F)** and **Paper (G)**).

The uncertainty analysis methodology follows the work of Sin et al. [20]: The Monte Carlo analysis of uncertainty involves three steps:

1. specifying the input uncertainty
2. sampling the input uncertainty
3. propagating the sampled input uncertainty in order to obtain a prediction uncertainty for the model output

A pre-exquisite of the application of the Monte Carlo based uncertainty analysis is that the model is specified containing process equations and property models.

2.3.1 Specification of property uncertainty and Monte Carlo sampling

First the input uncertainties for the property parameters need to be specified. These can be obtained from databases containing experimental measurement [15, 31, 32], GC based predictions (as described in section 2.1), or an estimation using Bootstrap method (as described in section 2.1).

A certain distribution needs to be assumed for the input parameter space of the properties (e.g. normal distribution with a standard deviation equal to the uncertainty range) and centred around the respective property value. These distributions are assumed to statistically represent the degree of belief with respect to where the appropriate values of the parameters lie.

The key step of the Monte Carlo procedure is the sampling of the parameter sets. The Latin Hypercube Sampling method [59] is utilized for probabilistic sampling of the fluid property input space of each compound. The procedure of Latin hypercube sampling is based on the division of the range of each property parameter in a certain number of equally proportioned intervals (principle of Latin square). A

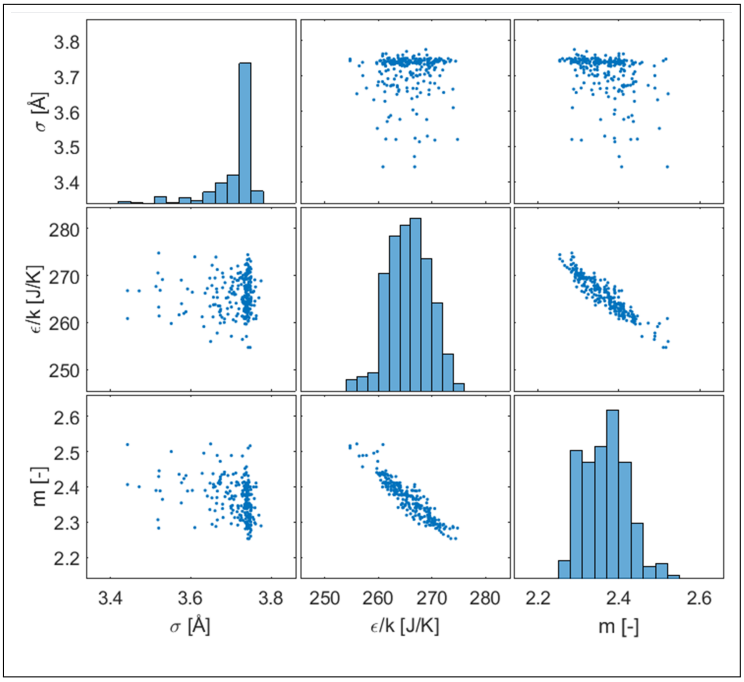


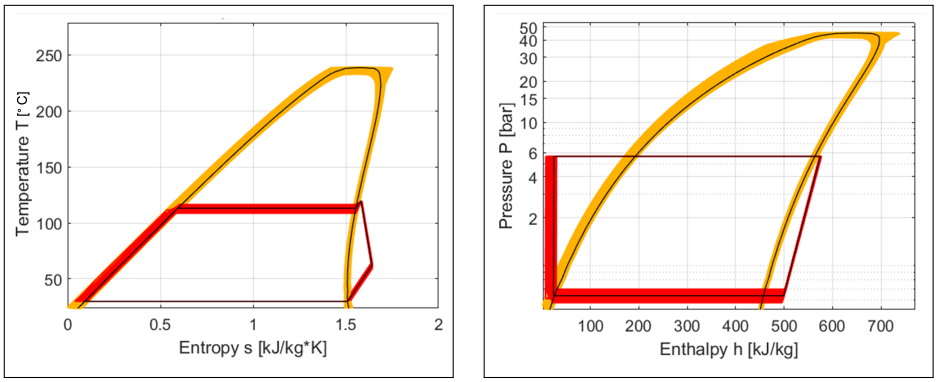
Figure 2.6: Illustration of samples generated by Latin Hypercube Sampling method with Iman and Conover correlation control for PC-SAFT parameters of cyclopentane. The distributions correspond to Figure 2.5.

sufficiently high number of samples (e.g. 200-400) is selected from the input parameter space, each sample containing one value for each input parameter. The sampling range is specified by the uncertainty (i.e. the 95%-confidence interval) range of each parameter.

The probability distribution of the parameters is taken into account. The rank-based method for correlation control of Iman and Conover [58] allows to take into account correlations between the input parameters. This is necessary, because property parameters (such as EoS parameters) are usually not completely independent.

As an example the Monte Carlo sampling procedure is performed for the PC-SAFT parameters of cyclopentane with the uncertainty range quantified by the bootstrap method (as shown in the previous section 2.2). The uncertainty specification and the correlation matrix can be found in Tables 2.1 and 2.2.

Figure 2.6 provides an illustration of the sampling results for the PC-SAFT parameters of cyclopentane.



(a) T-s-diagram. (b) log(P)-h diagram.

Figure 2.7: Representation of uncertainty with respect to the fluid properties in the T-s diagram and log(P)-h diagram for cyclopentane for SRK and PC-SAFT input uncertainty: Monte Carlo simulations overlaid (yellow/red) and mean (solid black line).

2.3.2 Model evaluation and quantification of model output uncertainty

One process model simulation is performed for each of the input parameter samples. The simulation results can provide a distribution of all the output variables of a thermodynamic cycle model (e.g. state variables, mass flow, etc.). As an example the PC-SAFT parameter samples of cyclopentane (shown in the previous section) are evaluated in an organic Rankine cycle (ORC) model (see **Paper (D)** for details).

The raw data obtained from the simulations can be plotted, for example in a temperature-entropy (T-s) or a logarithmic pressure-enthalpy (log(P)-h) diagram showing cycle points enumerated as in Figures 2.7. Each curve and design point set is different as different property parameter samples are used in each simulation. A varying band for both the saturation curves and the cycle design can be observed. From a statistical point of view the bands correspond to the distribution of the model outputs and directly visualize the sensitivity with respect to the fluid property values for the example compound cyclopentane.

For each of the simulations a certain cycle model output can be analysed separately. The distributions of the cycle model outputs are alternative representations of the uncertainty of the model with respect to the property parameters. Figure 2.8 depitcs as an example the net power outputs W_{NET} for the example of cyclopentane.

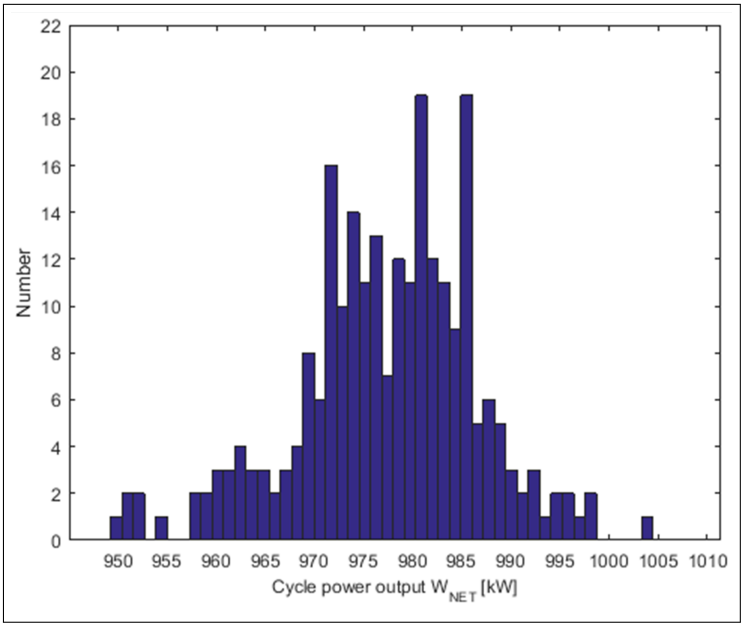


Figure 2.8: Distribution of the net power output of an ORC cycle (see **Paper (D)** with evaluated PC-SAFT samples of cyclopentane.

2.4 Reverse engineering of fluid design problem based on sampling and uncertainty analysis

Reverse engineering of a fluid design problem aims at identifying optimal fluid properties (giving a optimal virtual fluid) for a given process model and then subsequently find a real fluid in a database that matches this optimal properties [60]. Here a new approach for the reverse engineering of the working fluid design problem is presented. It is based on Monte Carlo sampling and uncertainty analysis. The detailed description including its application can be found in **Paper (G)**.

The methodology addresses in particular two challenges in fluid design:

- The difficulties in finding a feasible (real fluid) solution when solving a complex combined product-process design optimization problem (e.g. the identification of suitable working fluid candidates together with thermodynamic cycle process conditions).
- The incorporation of property uncertainties caused by measurements or property prediction.

The procedure of the new reverse engineering approach is divided in different steps:

1. Formulation of models for thermodynamic cycle and thermodynamic property estimation
2. Specification of working fluid property descriptors and search space for reverse engineering
3. Generation and evaluation of virtual fluids: Monte Carlo based sampling of property search space and evaluation in process model
4. Global sensitivity analysis of working fluid property descriptors and identification of weights for the property descriptors
5. Calculation of distance function between properties of real and virtual fluids and ranking of real fluids
6. Evaluation of identified high-ranked real working fluids including uncertainty analysis

Step 1: Formulation of models for thermodynamic cycle and thermodynamic property estimation

The thermodynamic cycle model and constraints have to be specified. The objective function of the cycle (e.g. maximum power output for an ORC or coefficient of performance for a heat pump) needs to be defined. Further, property models (e.g. EoS) have to be selected to predict the working fluid properties. The input properties to the EoS can be considered to be the major target properties, together with safety-related and environmental properties. For example if Peng-Robinson EoS is selected, the target properties to be identified for maximizing the objective function are: the critical temperature T_c, critical pressure P_c and the acentric factor ω. The goal is to identify a set of fluid target properties (corresponding to a real working fluid) that maximises the desired process objective function.

Step 2: Specification of working fluid property descriptors and search space for reverse engineering

For each of the target properties (property descriptors) a value range (lower and upper bound) needs to be specified. The ranges can be selected by analysing property data from a well-established database, e.g. DIPPR 801 AIChE database [31]. As an example the lower bound for the target property T_c can correspond to the 2.5% percentile of the DIPPR data set, whereas the upper bound corresponds to the 97.5% percentile. In this way very high or low values are excluded.

Step 3: Generation and evaluation of virtual fluids: Monte Carlo based sampling of property search space and evaluation in process model

This is the first key step of the new reverse engineering approach. Monte Carlo based sampling within the specified property search space is used to generate different sets of fluid-specific descriptors. These sets essentially can be considered as virtual fluids representing the search space. Low-discrepancy sampling is applied using uniformly distributed sequences (i.e. Halton sequences [61, 62]). This increases the uniform sampling of the property search space. Figure 2.9 provides an illustration of the sampling results for a search consisting of the Peng-Robinson EoS input parameters T_c, P_c and ω.

One model simulation in the cycle process model is performed for each of the virtual sample fluids. Subsequently the virtual fluids are ranked according to the highest cycle objective function (e.g. power output or coefficient of performance).

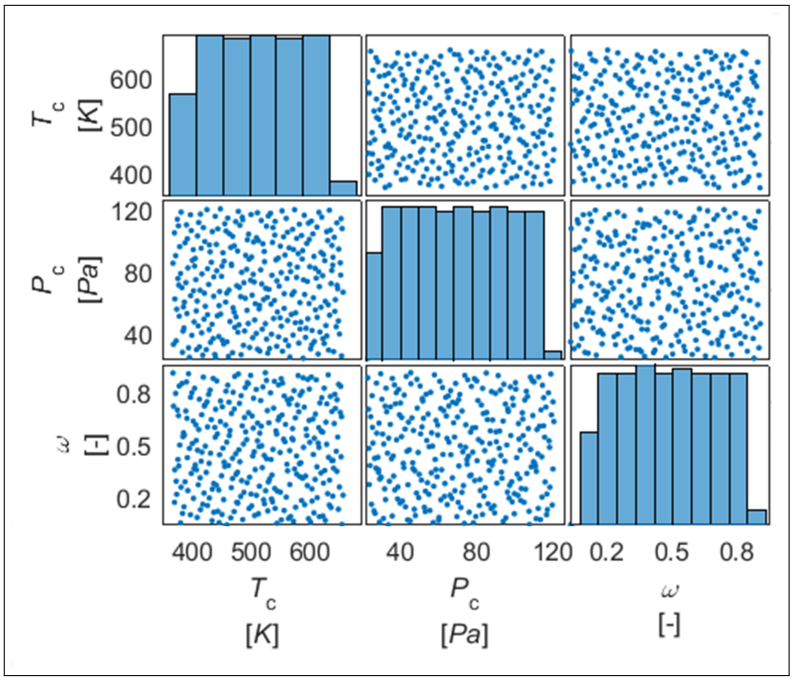


Figure 2.9: Illustration of samples generated by uniform low-discrepancy sampling using Halton sequences [61]: The sample matrix consisting of (200-400 samples) represents the property search space. The diagonal elements of the matrix represent the uniform distribution of the sampling.

Step 4: Global sensitivity analysis of working fluid property descriptors and identification of property weights

A global sensitivity analysis of the target properties with respect to the thermodynamic model output (e.g. power output or coefficient of performance) is used to analyse the impact of the respective parameters in the thermodynamic cycle model [63]. Property parameters (target properties) can have a high or low sensitivity with respect to the cycle model output. If they have a high sensitivity, the respective property descriptor should have a higher weight (importance factor), when identifying the best suitable working fluid from the database.

The derivative-based global sensitivity measures is used following the work of Kucherenko et al. [64] to investigate the overall influence of the property descriptors in the search space. This method provides so called global sensitivity measures of the corresponding target properties. Afterwards, a normalized weight factor w can be assigned to the respective target property. The details of the global sensitivity analysis methods can be found in **Paper (G)**. Furthermore, **Paper (H)** discusses the use of global sensitivity analysis tools for the ranking of target properties.

Step 5: Calculation of distance function between properties of real and virtual fluids and ranking of real fluids

This is the second key step of the new approach. The goal is to identify real working fluids from a chemical database that are closest to the top performing virtual (sample) fluid.

First, a database containing property data of a large amount of real chemical compounds needs to be selected, e.g. NIST ThermoData Engine (TDE) [32, 33] or DECHEMA DETHERM [34].

The distances between the property value of a top performing virtual fluid and a real fluid is calculated including the property uncertainty range of the respective real fluid. Well-established databases, such as TDE, provide every property value with its corresponding uncertainty range. Alternatively, the uncertainty range can be determined by the measurement [15] or it can be obtained using GC based prediction (as described in section 2.1).

As an example the calculation of the distances of the target property acentric factor ω is shown. The acentric factor of a real fluid y is considered as

$$\omega_y^{\text{low}} < \omega_y < \omega_y^{\text{up}} \tag{2.13}$$

where ω_y is the value of the acentric factor, ω_y^{low} is the lower bound of the uncertainty range for the database value and ω_y^{up} is the upper bound respectively.

The distance between a top performing virtual fluid x and a real fluid y in the property search space of the acentric factor ω is defined as d_{xy}^{ω}. The distance between the virtual fluid and the real fluid is calculated from the uncertainty range. This is not done in classical reverse engineering approaches as described in the literature [60, 65], where the distance is usually calculated from the property value itself, not from the uncertainty bound.

For example, if the acentric factor ω_x of the top performing virtual fluid x lies within the uncertainty range of the acentric factor ω_y of the real fluid y, the distance function is assigned a zero value. However, if the virtual property ω_x is below the lower bound ω_y^{low} of the real property, then the normalized distance between ω_x and ω_y is calculated as follows

$$\text{if} \quad \omega_x < \omega_y^{\text{low}} \quad \text{then} \quad d_{xy}^{\omega} = \frac{|\omega_y^{\text{low}} - \omega_x|}{\omega_y^{\text{low}}} \tag{2.14}$$

where $|\omega_y^{\text{low}} - \omega_x|$ corresponds to the absolute value norm, i.e. L1 norm [66].

Figure 2.10 illustrates the principle for the calculation of the distance function for the two-dimensional search space of ω and T_c.

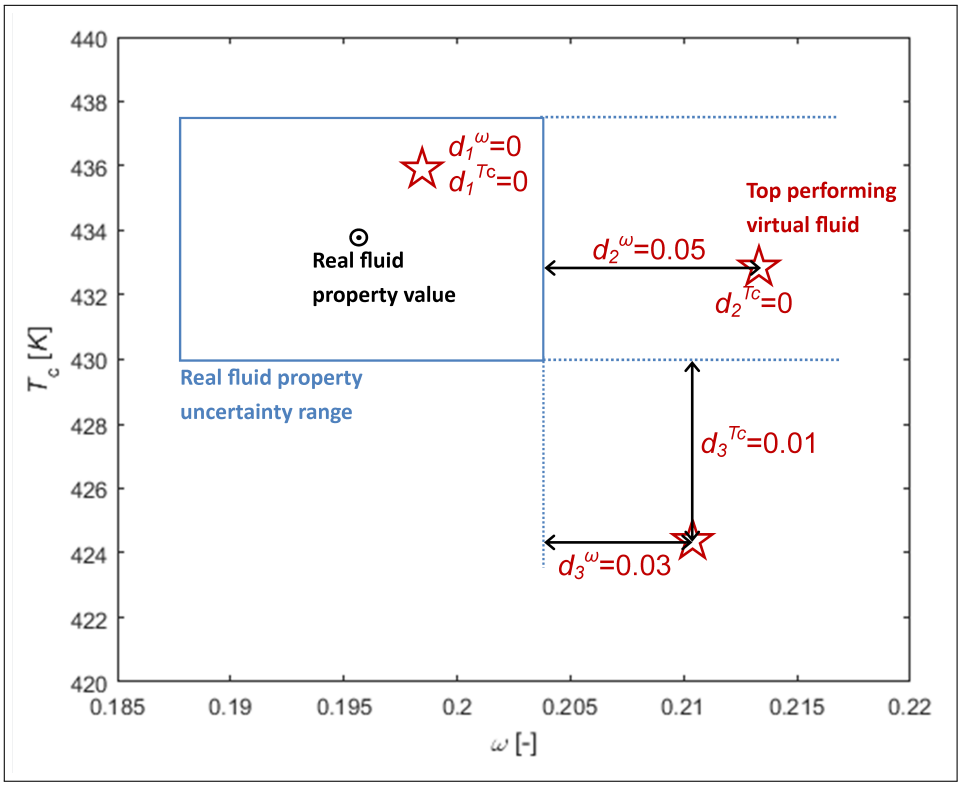


Figure 2.10: Illustration of the algorithm to calculate the distance function in the ω-T_c sub-search space: If the virtual fluid (stars) value lies within the uncertainty bounds of the respective real fluids (circle), the distance d_{xy} is zero. Otherwise the distance is calculated between the uncertainty bounds and the virtual fluid.

The total distance function d_{xy}^{tot} for the distance of one virtual fluid x to one real fluid y is calculated by summing up all the property distances multiplied by its corresponding weight factors w (obtained from Step 4). For the example of three target properties (T_c, P_c and ω) the total distance is given by

$$d_{xy}^{\text{tot}} = d_{xy}^{\omega} \cdot w^{\omega} + d_{xy}^{Tc} \cdot w^{Tc} + d_{xy}^{Pc} \cdot w^{Pc} \tag{2.15}$$

The distance function can be written in a general form for N properties i as

$$d_{xy}^{\text{tot}} = \sum_{i}^{N} d_{xy}^{i} \cdot w^{i} \tag{2.16}$$

Afterwards the real working fluids are ranked according to the lowest total distance function value d_{xy}^{tot}.

Step 6: Evaluation of identified high-ranked real working fluids including uncertainty analysis

The high ranked real working fluids are evaluated in the cycle model and the Monte Carlo based uncertainty propagation (as described in section 2.3) with respect to the property value uncertainties is performed for each of the compounds. This provides a cumulative distribution function for the model output (e.g. net power or coefficient of performance) of each of the real fluids (as shown in Figure 2.8).

The analysis of the cycle output uncertainty for each real fluid allows to rank the promising fluid candidates according to the mean value, the lower bound (conservative approach) or the upper bound (optimistic approach) of the respective cycle objective function. Hence, property uncertainty can be considered as a criterion when selecting a promising fluid.

Chapter 3

Applications and case studies

The methodologies and tools described in the previous chapter are applied to different applications and case studies, which are outlined here. The property uncertainty analysis tools are applied to the development of a new prediction method for safety-related properties and further for the comparison and assessment of three equations of state based on their uncertainties. In order to show the identification of optimal working fluids using the methodologies described in the previous chapter, two industrial case studies are described: 1) an organic Rankine cycle for the recovery of low-temperature heat from the exhaust gas of a marine diesel engine, and 2) a heat pump system for low-temperature heat recovery in food industries. All the applications and case studies have been implemented and simulated in the numerical computing software *MATLAB* [67].

3.1 Property prediction and uncertainty analysis

3.1.1 New prediction methods for safety-related properties

A new set of improved group contribution (GC) models for safety-related properties is developed through the application of the systematic model development and analysis method, which is described in section 2.1. The models include a thorough uncertainty analysis (i.e. estimation of the 95%-confidence interval) of every prediction. The details of the application can be found in **Papers (A)**, **(B)** and **(C)** in Part II.

The safety characteristics of hazardous substances provide indispensable information for the risk assessment of chemical products in industrial and domestic processes. In particular flammability-related properties such as the heat of combustion ΔH_c°, the lower and upper flammability limit (LFL and UFL), the flash point (FP) and the auto ignition temperature (AIT) are important to quantify the risk of fire and explosion [68].

LFL and UFL are defined as the lowest and the highest possible concentration of a substance in air at which a flammable mixture is formed. ΔH_c° is defined as the enthalpy increase of a chemical compound, while undergoing an oxidation to defined combustion products [69]. LFL, UFL, and ΔH_c° are stated at a specific temperature (298.15 K) and pressure (1 atm). However, LFL and UFL change with increasing temperature [70]. FP is the lowest temperature, where a liquid forms an ignitable vapour-air mixture. AIT is the lowest possible temperature, above which a substance will ignite in air without an external ignition source [68].

Natural refrigerants, such as hydrocarbons, show promising performance in thermodynamic cycles, have no ozone depletion potential, and possess much lower global warming potential compared to fluorinated and chlorinated compounds, some of which are being phased-out in Europe [13]. However, the disadvantage of hydrocarbon based working fluids is that many are highly flammable [7]. Whenever experimental values on safety-related properties are unavailable (e.g. in the early design phase), flammability property prediction models become a crucial tool to estimate the hazard associated with a fluid.

The procedure (described in section 2.1) is applied to develop the GC models for the single point LFL, UFL, ΔH_c°, FP and AIT, to estimate its parameters and to perform the uncertainty analysis. The GC factors for FP and AIT have already been described by Hukkerikar et al. [16] and are re-estimated using robust regression and outlier treatment, aiming at an improved parameter fit compared to the previous estimations.

In Eq. (3.1) to (3.1.1) functions are suggested for ΔH_c°, LFL, UFL and FP

for a specific compound. They are obtained by plotting various classes of pure components versus their increasing carbon number in homologue series, in order to obtain ideas regarding the property function $f(X)$ (see also Figure 2.3 in section 2.1).

$$f(\Delta H_c^\circ) = \Delta H_c^\circ - \Delta H_{c\,const}^\circ = \sum_j N_j^{\Delta H} C_j^{\Delta H} + \sum_k M_k^{\Delta H} D_k^{\Delta H} + \sum_l O_l^{\Delta H} E_l^{\Delta H} \tag{3.1}$$

$$f(LFL) = log\left(\frac{LFL}{LFL_{const}}\right) = \sum_j N_j^{LFL} C_j^{LFL} + \sum_k M_k^{LFL} D_k^{LFL} + \sum_l O_l^{LFL} E_l^{LFL} \tag{3.2}$$

$$f(UFL) = log\left(\frac{UFL}{UFL_{const}}\right) = \sum_j N_j^{UFL} C_j^{UFL} + \sum_k M_k^{UFL} D_k^{UFL} + \sum_l O_l^{UFL} E_l^{UFL} \tag{3.3}$$

$$f(FP) = FP - FP_{const} = \sum_j N_j^{FP} C_j^{FP} + \sum_k M_k^{FP} D_k^{FP} + \sum_l O_l^{FP} E_l^{FP} \tag{3.4}$$

$$\begin{aligned} f(AIT) = AIT_{const1} + \left(\sum_j P_j^{AIT} X_j^{AIT} + \sum_k Q_k^{AIT} Y_k^{AIT} + \sum_l R_l^{AIT} Z_l^{AIT}\right) \\ + AIT_{const2} \cdot 10^{-\left(\sum_j N_j^{AIT} C_j^{AIT} + \sum_k M_k^{AIT} D_k^{AIT} + \sum_l O_l^{AIT} E_l^{AIT}\right)} \end{aligned} \tag{3.5}$$

The sum on the right hand side corresponds to the Marrero/Gani (MG) GC model structure as described in section 2.1. $\Delta H_{c\,const}^\circ$, LFL_{const}, UFL_{const}, FP_{const}, AIT_{const1}, AIT_{const2} are universal constants that need to be determined by the parameter regression.

In order to account for the temperature-dependence of LFL the approach of Rowley et al. [28] is used as a basis to derive a new MG GC method. The detailed derivation can be found in **Paper (C)**. The temperature-dependence of LFL for compound i is described as

$$LFL_i(T) = LFL_i(T_{ref}) + K_i^{LFL} \cdot (T - T_{ref}) \tag{3.6}$$

where K_i^{LFL} is the proportionality constant between LFL and T for a specific compound i. In this application it is proposed to estimate K_i^{LFL} by a MG GC model:

$$f(K^{LFL}) = \frac{K_{const}^{LFL}}{K^{LFL}} = \sum_j N_j C_j \tag{3.7}$$

with K_{const}^{LFL} as the universal correlation constant and C_j the first order parameters that occurs N_j times.

Experimental data for ΔH_c°, LFL, UFL, FP and AIT are taken from AIChE DIPPR 801 Database [31]. Only data points that are classified by DIPPR as “experimental”

and “accepted” values are considered. Data for the temperature-dependence of LFL are collected from different sources [70, 71, 72].

An overview of the results of the application of the described framework on the safety-related properties described above is shown in section 4.1. A detailed analysis of the results is given in **Papers (A)**, **(B)** and **(C)** in Part II.

3.1.2 Comparison of equations of state: SRK, Peng-Robinson and PC-SAFT

The second application on uncertainty analysis focuses on the uncertainty assessment of equations of state in the context of an organic Rankine Cycle (ORC). The study is also described in **Paper (D)** and **Paper (E)** in part II.

The aim is to compare the uncertainties from both the mathematical form of the EoS and from the data for obtaining the EoS parameter values. Furthermore, a statement should be made, which EoS performs better in terms of uncertaity on the cycle level. Hence, uncertainty is presented as an assessment criterion of EoS (together with e.g. goodness-of-fits to data or limited complexity [73]). The considered cycle is an industrial ORC for the recovery of exhaust gas heat from a marine diesel engine vessel [6]. Two cubic EoS, SRK and Peng-Robinson, are compared to the non-associating PC-SAFT EoS. These EoS only require three fluid-specific input properties and have been extensively used for screening or molecular design studies of working fluids for thermodynamic cycles [74]. However, their uncertainty propagation through a thermodynamic cycle has not been addressed in the literature.

For the study the natural refrigerant cyclopentane is selected. The overall methodology consists of the major steps given in Table 3.1.

Table 3.1: Overview of the applied methodology.

Step 1	Formulations of EoS and fluid selection
Step 2	Organic Rankine cycle (ORC) process model formulation and optimization of process parameters
Step 3	Quantification of uncertainties in fluid-specific EoS parameters based on experimental data for cubic SRK, Peng-Robinson and PC-SAFT EoS (following section 2.2 of the framework)
Step 4	Monte Carlo procedure for input uncertainty propagation to ORC process model output of cubic SRK, Peng-Robinson and PC-SAFT EoS (following section 2.3 of the framework)
Step 5	Comparison of EoS with respect to ORC output uncertainty

Step 1: Formulations of EoS and fluid selection

The methodology is applied to the following equations and parametrizations, which will be briefly outlined:

- non-associating PC-SAFT [55] EoS pametrized in σ, ε/k and m

- cubic Peng-Robinson [53] and SRK [54] EoS pametrized in T_c, P_c and ω
- cubic Peng-Robinson and SRK EoS pametrized in a, b and β
- Aly-Lee ideal gas heat capacity [75] parametrized A, B, C, D, E (for the ideal gas contribution of the EoS)

PC-SAFT is based on a statistical thermodynamic theory for fluids with a repulsive core and directional short-range attractive sites. The non-ideal Helmholtz energy, A_{res}, of a system chain molecules has the form

$$A_{res} = A_{hc} + A_{disp} \tag{3.8}$$

with A_{hc} being the hard-chain reference contribution and A_{disp} being the dispersion contribution. Details of the derivation, structure and the thermodynamic properties of PC-SAFT can be found in the work of Gross et al. [55]. In general, the PC-SAFT EoS is parametrized in terms of the parameters σ (segment diameter), ε/k (energy parameter), and m (chain length parameter).

The uncertainties of the PC-SAFT parameters σ, ε/k, and m are obtained through fitting to collected experimental data for vapor pressure [56] over the temperature range of 230-350K and saturated liquid densities [57] for a temperature range of 190-310K using a the Bootstrap method (as described in section 2.2). The uncertainties in σ, ε/k, and m are afterwards propagated through an ORC model system to obtain the uncertainty of the ORC model outputs (i.e., the net power output uncertainty).

The three-parameter cubic EoS can be written in the general form

$$P = \frac{RT}{V_m - b} - \frac{a\alpha(T)}{V_m^2 + (c+1)\cdot V_m \cdot b - c\cdot b^2} \tag{3.9}$$

with T being the absolute temperature, P the absolute pressure, V_m the molar volume and R the universal gas constant. The parameters a, b and c as well as the temperature-dependent function $\alpha(T)$ are specific for the particular version of the cubic EoS (i.e. SRK or Peng-Robinson)[14].

The SRK and Peng-Robinson EoS input property parameters as recommended in the literature are critical temperature, T_c, critical pressure, P_c, and acentric factor, ω, so as to ensure the inflection of the critical isotherm at the critical pressure [76] and to (nearly) reproduce the vapor pressure used to obtain the acentric factor. For T_c measurement uncertainty [14, 77] serves as input uncertainty to the EoS. P_c and ω are obtained using the Bootstrap method to fit vapor pressure to an Antoine equation as described by Patel and Ambrose [78, 14].

Additionally, the cubic EoS are fit to the same experimental data as used previously for the PC-SAFT EoS and its uncertainty is quantified using a Bootstrap

method. This allows comparison of uncertainties in the data fitting between cubic and PC-SAFT EoS. To do this, it is required to re-parameterize the two cubic EoS in terms of their parameters a and b as well as a third parameter β. For SRK the re-parameterized $\alpha(T)$-function of the cubic EoS is given by

$$\alpha(T) = \left(1 + \beta \cdot \left(1 - \sqrt{\frac{T}{T_c}}\right)\right)^2 \tag{3.10}$$

A similar formulation is used for Peng-Robinson EoS. In this way the isotherm is not ensured with an inflection to be at the critical temperature; instead a, b and β are considered as fluid-specific parameters that should be obtained by fitting the EoS to experimental data. This allows to fit the SRK and Peng-Robinson EoS to the same data as PC-SAFT EoS.

The computational coding of the PR and SRK EoS is adapted from the work of Liu et. al [79], whereas the PC-SAFT EoS implementation is based on the work of Gross et al. [55] and of Fakouri Baygi et al. [80].

The ideal-gas enthalpy and entropy terms are obtained by integrating a temperature-dependent ideal-gas heat capacity function, $c_p(T)$, with parameters obtained from fitting thermal or spectroscopic measurements combined with molecular theory. The Aly-Lee form of the ideal-gas heat capacity with five compound-specific input parameters (A, B, C, D, E) [75] is used:

$$c_p(T) = A + B \cdot \left[\frac{\frac{C}{T}}{\sinh\left(\frac{C}{T}\right)}\right]^2 + D \cdot \left[\frac{\frac{E}{T}}{\cosh\left(\frac{E}{T}\right)}\right]^2 \tag{3.11}$$

Step 2: Organic Rankine cycle (ORC) process model formulation and optimization of process parameters

The quantified parameter uncertainties of the corresponding EoS are propagated through an industrial ORC application for power generation using a low-temperature heat source. The ORC is designed to recover heat from exhaust gas of a marine diesel engine of a large container ship vessel [81].

Cyclopentane is the working fluid. The process model is based on the work of Andreasen et al. [82]. The detailed model description and equations can be found in **Paper (D)** as well as **Paper (E)** in Part II. Figure 3.1 gives an overview over the system containing the components and the corresponding modelling constraints of the process and of the hot fluid.

The process data is provided by MAN Diesel and Turbo [6]. The modelling constraints of the process and of the hot fluid are summarized in Figure 3.1.

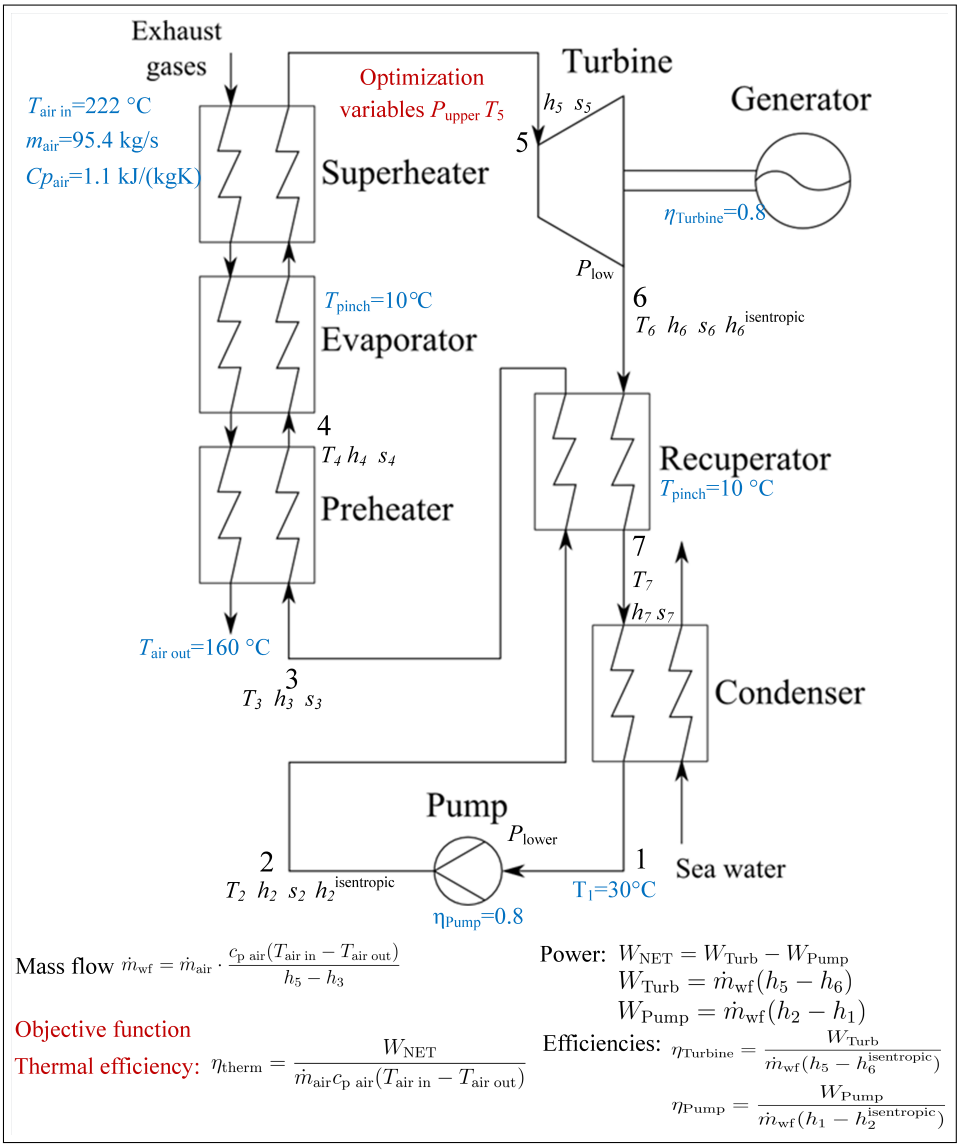


Figure 3.1: An overview over the ORC process adapted from Andreasen et al. [82]. The objective function is the thermal efficiency η_{therm}, which is optimized subject to the objective variables P_{upper} and T_5 (red) and the specified process parameters (blue). The outputs from the ORC process model are the net power output W_{NET}, the mass flow $\dot{m}_{wf}$ of the working fluid, and state variables such as pressures P_i, temperatures T_i, entropies s_i, and enthalpies h_i. The specified process variables are the exhaust gas input/ouput temperature, mass flow and heat capacity ($T_{\text{air in}}$, $T_{\text{air out}}$ m_{air} $C_{p\ \text{air}}$), the condensation temperature (T_1), the pinch temperature (T_{pinch}) and the turbine and pump efficiencies (η_{Turbine}, η_{Pump}).

According to a degrees-of-freedom analysis of the cycle, two process variables can be solved for and optimized. The turbine inlet pressure, $P_5 = P_{\text{upper}}$, and temperature, T_5 are selected. The optimal process conditions are identified by performing a particle swarm optimization for cyclopentane [83].

Step 3: Quantification of uncertainties in fluid-specific EoS parameters based on experimental data for cubic SRK, Peng-Robinson and PC-SAFT EoS

The quatification of the uncertainties of the fluid-specific EoS parameters is performed through the Boostrap method as outlined in section 2.2 of the framework. The details of the calculations can be found in **Paper (E)**.

Step 4: Monte Carlo procedure for input uncertainty propagation to process model output of cubic SRK, Peng-Robinson and PC-SAFT EoS

For each of the EoS the Monte Carlo based procedure is performed to propagate the EoS parameter uncertainties through the process model onto the ORC cycle output. The methodology is part of the overall framework of this thesis and is described in section 2.3. As a result distributions of the cycle output variables (e.g. W_{NET}) are obtained, that represent the output uncertainties subject to the quantified property uncertainties.

Step 5: Comparison of EoS with respect to ORC output uncertainty

The results of the Monte Carlo uncertainty propagations are distributions of the model outputs (e.g. the net power output W_{NET} of the ORC). The broader a model output distribution is, the more uncertain is the model output value. Based on the distributions of the model outputs from the Monte Carlo simulations, the following questions can be addressed:

1. Do input uncertainties originating from the ideal-gas contribution or from the departure functions have a stronger influence on the model output?
2. Which of the departure function input uncertainties (SRK, Peng-Robinson or PC-SAFT) has the strongest effect on the model output?
3. What differences occur for the uncertainties related to different parametrizations of cubic EoS?
4. Is the uncertainty propagation from properties to process outcomes more determined by the model form or by the data used in regressions?

The results of the study are summarized in section 4.2. A detailed analysis of the results is given in **Papers (D)** and **(E)** in Part II.

3.2 Working fluid selection

3.2.1 Organic Rankine cycle for low-temperature heat recovery

This case study aims to select working fluids for an Organic Rankine Cycles (ORC) taking into account property uncertainties of the working fluids. The detailed description of the selection procedure is given in **Paper (F)**. This is the first case study that systematically shows how to implement a property-focused uncertainty analysis into the working fluid selection problem and includes specifically the uncertainty information as additional quantitative criterion for the fluid selection.

An overview of the methodology applied to the case study and divided in different steps is shown in Table 3.2.

Table 3.2: Overview of the applied methodology.

Step 1	Formulation of ORC process models and constraints
Step 2	Selection of EoS and property database
Step 3	Model solution for all DIPPR database compounds and identification of optimal process variables
Step 4	Monte Carlo procedure for input uncertainty propagation to ORC process model output for all feasible compounds (following section 2.3 of the framework)
Step 5	Ranking of the fluids including uncertainty

Step 1: Formulation of ORC process models and constraints

The ORC process investigated in this study is sketched in Figure 3.2 and is based on the work of Andreasen et al. [82]. It is a general ORC for waste heat recovery from a hot process water stream. The layout consists of four main components: a pump, a boiler (preheater, evaporator and superheater), a turbine and a condenser. The working fluid is an organic compound, which is circulated by the pump. The base case fluid is 1,1,1,3,3-Pentafluoropropane (R-245fa or HFC-245fa).

The hot fluid is water at a temperature of 120°C and a mass flow of 50 kg/s, representative of a waste heat stream of a chemical plant or a geothermal heat source [82]. There are no limitations imposed on the hot fluid outlet temperature. Two process variables can be solved for and optimized. The turbine inlet pressure, P_t, and the turbine inlet temperature, T_t, are selected.

All modelling assumptions are described in detail in **Paper (F)** and all the process equations are provided in the appendix of **Paper (F)** in Part II.

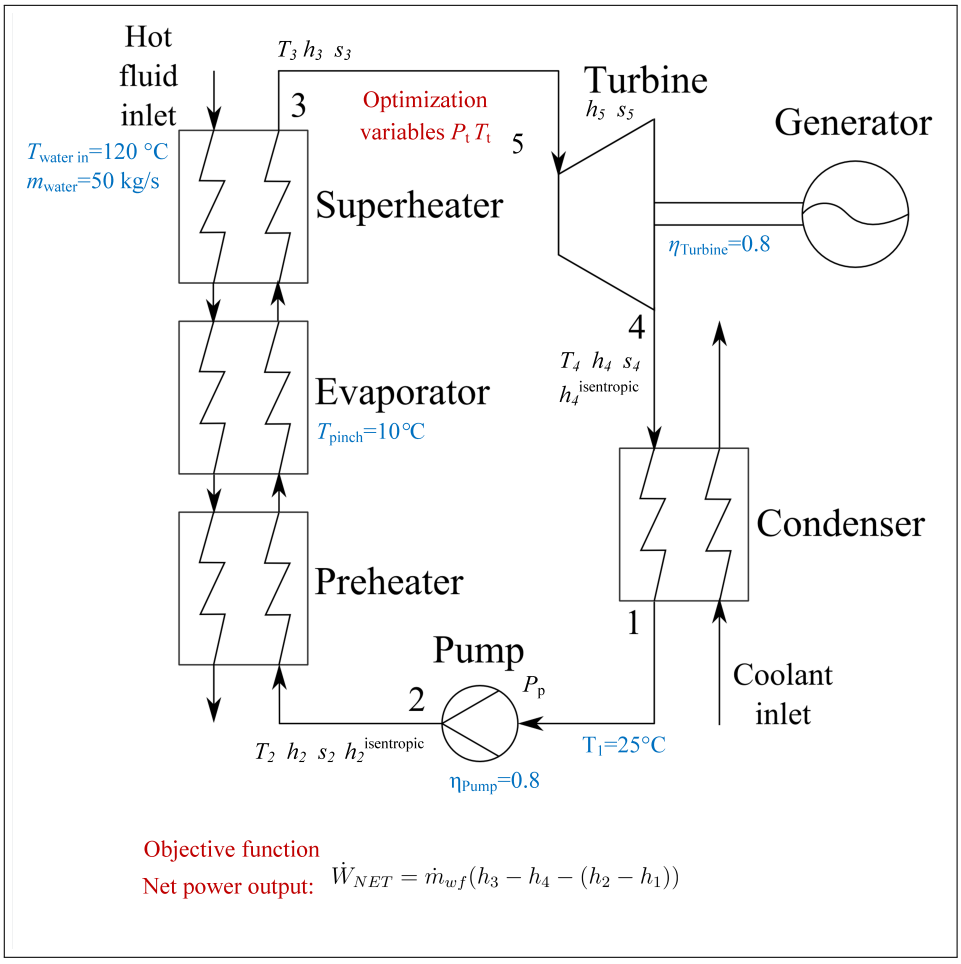


Figure 3.2: An overview over the ORC process [82] (see also **Paper (F)** in Part II). The objective function is the net power output W_{NET}, which is optimized subject to the objective variables P_{t} and T_{t} (red) and the specified process parameters (blue). The specified process variables are the exhaust gas input temperature and mass flow ($T_{\mathrm{water\ in}}$, m_{water}), the condensation temperature (T_1), the pinch temperature (T_{pinch}), the turbine and pump efficiencies (η_{Turbine}, η_{Pump}).

Step 2: Selection of EoS and property database

Peng-Robinson EoS [84] parametrized in T_{c}, P_{c} and ω is selected in order to determine the departure functions of the thermodynamic properties, because of its relatively small number of required fundamental parameters as a cubic equation of state, which makes it suitable for the screening of a large number of possible working fluid candidates [79]. Ideal gas enthalpy and entropy changes are calculated by integrating the temperature-dependent ideal gas heat capacity as defined by Aly

and Lee [75].

The fluid parameter inputs for Peng-Robinson EoS are the molecular weight MW, the critical temperature T_c, critical pressure P_c, and the acentric factor ω. Therefore, the evaluation of the thermodynamic properties required for the ORC model needs only four primary fluid properties (MW, T_c, P_c and ω) and the respective Ally-Lee heat capacity constants (A, B, C, D, E).

All these properties can be found in the DIPPR 808 AIChE database [31] for 1965 chemical compounds. The database values for (T_c, P_c and ω) can be both experimental and predicted. DIPPR provides the Ally-Lee heat capacity constants that have been obtained by fitting the Ally-Lee correlations for each substance to the respective experimental or predicted temperature dependent heat capacity curve. The DIPPR database states the respective uncertainty of T_c, P_c and ω along with the heat capacity values obtained from the constants A, B, C, D and E [85]. The information on the uncertainty is further used to calculate the output uncertainty of the net power.

The goal of the ORC model is to identify the working fluid with its corresponding optimal process parameters that provides the highest net power output.

The ORC modelling results with Peng-Robinson EoS are compared to the results obtained with REFPROP 9.0 [51] for those compounds for which REFPROP parameters are known. The values obtained with REFPROP have (with two exceptions) a relative deviation below 2%, implying that the numerical models give reasonable results.

Step 3: Model solution for all DIPPR database compounds and identification of optimal process variables

For each of the 1965 chemical compounds from the DIPPR 801 AIChE database, the optimal process variables (turbine input pressure P_t and temperature T_t) are identified and the corresponding net power output needs to be calculated. An alternative to the usage of an optimization algorithm is used here to identify the optimal process conditions (in contrast to application 3.1.2, the uncertainty assessment of EoS).

Monte Carlo based sampling procedure is utilized to obtain a number of 250 uniformly distributed pairs of process variables P_t and T_t. The values are sampled within the predefined variable constraints in a temperature range between 25 and 110°C (corresponding to the condensation temperature and the heat source temperature) and a pressure range between 1 and 15 bar. For all of the compounds the ORC model is evaluated using the sampled pairs of process variables. Then, the process variables giving the highest net power output are chosen to be the most

favorable quasi-optimal pair of process variables (P_t and T_t) for the respective compound. Comparison of particle swarm optimization [83] with the sampling based optimization approach used here showed an average average relative deviation of 1.01% for W_{NET}, 0.40 % for P_t and 1.01% for T_t.

Step 4: Monte Carlo procedure for input uncertainty propagation to ORC process model output for all feasible compounds

For each of the feasible chemical compounds obtained from Step 3 the Monte Carlo based uncertainty analysis focusing on the input property uncertainty is performed. The Monte Carlo based procedure follows the methodology outlined in section 2.3. The Monte Carlo results provide a distribution for the net power output of each compound representing input uncertainties propagated through the cycle model. The 95%-confidence interval of the net power output can be calculated.

Step 5: Ranking of the fluids including uncertainty

The compounds are ranked according to their respective net power output including the 95%-confidence interval. This enables an assessment of the compounds not solely based on their actual cycle performance, but also according to the reliability of the property data used.

The case study results are presented in section 5.1. The details of the results are shown in **Papers (F)** in Part II.

3.2.2 Heat pump system in food industries

In this case study the newly developed sampling based reverse engineering approach of the working fluid design problem is applied. The detailed description of the reverse engineering application is provided in **Paper (G)**. This case study also took advantage of two other methodologies described in this thesis: The Monte Carlo method for property uncertainty propagation (section 2.3) and the developed GC based prediction method for the lower flammability limit (section 3.1.1).

The procedure described in section 2.4 is followed: 1) Problem formulation 2) Specification of property range 3) Generation of virtual fluids using Monte Carlo sampling 4) Global senstivitity analysis of working fluid property 5) Calculation of distance function between real and virtual fluids 6) Evaluation and ranking of fluids including uncertainty analysis.

In this study optimal pure component working fluid candidates for an industrial heat pump system are to be identified. The heat pump of this study is a waste heat recovery system used in a spray drying facility for milk powder production in a reference dairy factory located in Denmark [2]. The model is based in the work of Zühlsdorf et al. [86]. The heat pump system is used to recover heat from the outlet gas of a spray dryer (heat source) and preheat the air before the spray dryer (heat sink), utilizing secondary cycles with pressurized water. The heat pump is outlined in Figure 3.3, including the constraints concerning the conditions used for the hot fluid and the process components.

Heat is provided by the water source to the working fluid through the evaporator (state 6 to 7 in Figure 3.3 and the super-heater (state 7 to 1). The heat recovery performed by the system should heat up the sink water from 75°C to 125°C.

A degree of freedom analysis shows that the cycle can be solved by optimizing two process variables, which are chosen to be the condensation and the evaporation pressure. The optimal process conditions for each set of property descriptors (i.e. virtual fluids) are identified by Newton–Raphson method [87] with the coefficient of performance (*COP*) as the objective function. The optimization problem in this study is much simpler compared to other studies, that solved e.g. MINLP problems [88].

Peng-Robinson EoS [53] parametrized in T_c, P_c and ω is selected to determine the departure functions of the thermodynamic properties. For the EoS ideal gas contribution, a linear temperature dependence of the isobaric ideal gas heat capacity (expressed by two parameters A and B only) is used in order to lower the amount of fluid specific parameters for the reverse engineering problem [60]: $c_p(T) = A + B \cdot T$.

The fluid-specific property descriptors [89] for the reverse engineering problem need to be specified. For the given case study the property input parameters to the

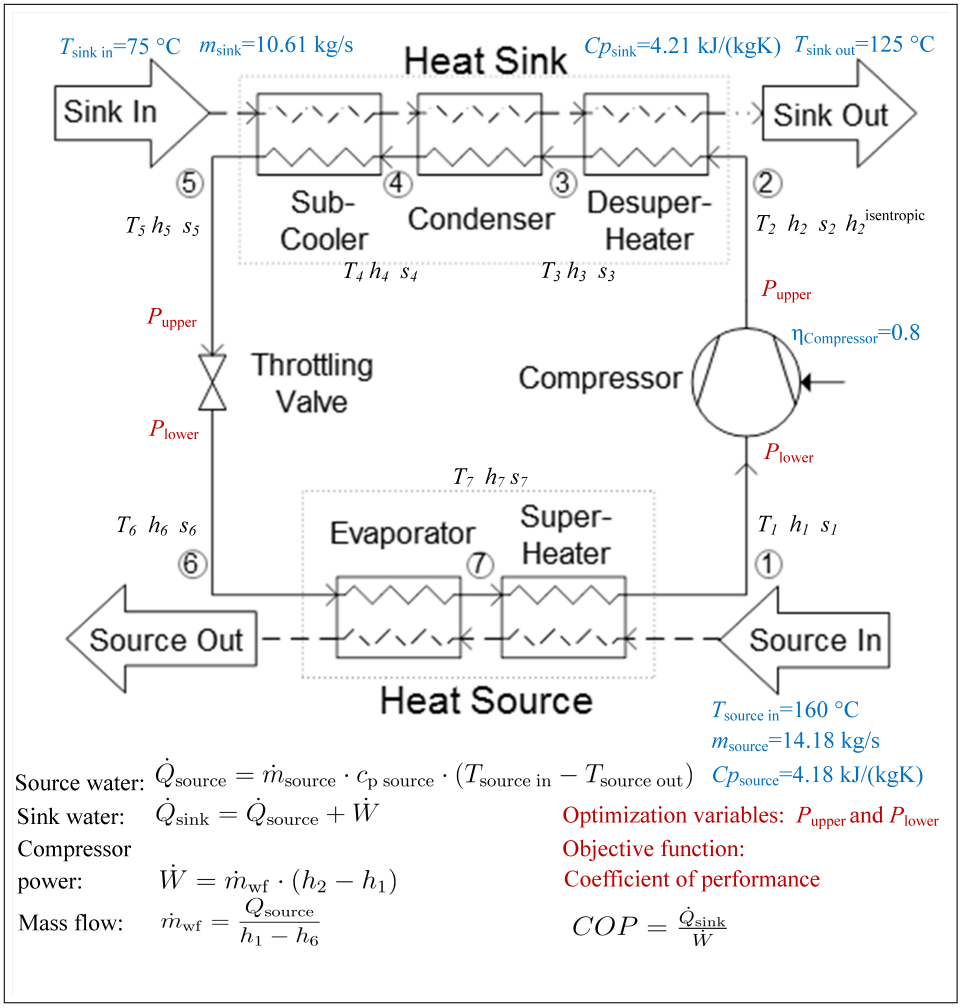


Figure 3.3: An outline of the industrial heat pump for waste heat recovery [86] (see also **Paper (G)** in Part II). The outputs of the heat pump model are the coefficient of performance *COP* (i.e. the ratio between the supplied heat and the compressor power input to the system), the working fluid mass flow $\dot{m}_{wf}$, and state variables, temperatures T_i, entropies s_i, and enthalpies h_i. $\dot{Q}_{source}$ is the source heat, $\dot{Q}_{sink}$ is the heat provided to the water sink and $\dot{W}$ is the compressor power. $\dot{m}_{\text{source}}$ expresses the source water mass flow, $c_{\text{p source}}$ and $c_{\text{p sink}}$ are the corresponding water heat capacities. $T_{\text{source in}}$ and $T_{\text{source out}}$ are the source input and output temperatures. The optimization variables are the upper and lower pressure, P_{upper} and P_{lower}.

ideal-gas contribution and departure function of the EoS are chosen. These are:

- heat capacity correlation constants: A and B
- critical temperature T_c

- critical pressure P_c
- acentric factor ω
- molecular weight MW

For each of the six property descriptors a value range (lower and upper bound) is specified by analysing property data from the well-established DIPPR 801 AIChE database [31].

Afterwards virtual sample fluids are generated by Monte Carlo sampling and evaluated as described in Step 3 of the reverse engineering methodology (described in section 2.4). The global sensitivity analysis of the properties allows assigning weights for the target properties (Step 4).

In the current case study the focus lies on working fluids that are based on cyclic hydrocarbons. Hence, the new methodology should identify novel pure component cyclic hydrocarbons as working fluids. However, the methodology itself is generic and can be applied to screen larger classes of chemicals database, the latter focus is made intentional so as to remain focused on demonstrating a proof of concept of the new methodology. In total 2126 real cyclic hydrocarbon based compounds from TDE database [32, 33] with property data including uncertainty are used as real fluids for calculating the distance to the best performing virtual fluids (Step 5).

In the context of the phase-out of chlorinated and fluorinated compounds for thermodynamic cycles in Europe [12], hydrocarbon based (natural) refrigerants show promising performances have no ozone depletion potential, and possess much lower global warming potential [13]. The TDE database contains a large number of cyclic hydrocarbons. Fluids of such type (e.g. cyclopentane) are often not considered when performing fluid design with "classical" computer-aided molecular design optimization algorithms due to the high number combinatorial possibilities and the difficulties of estimating property data for such compounds [88]. Hence, cyclic hydrocarbons have not been extensively considered as hydrocarbon based refrigerants in the literature when performing fluid design with molecular design optimization algorithms.

In the last step of the methodology a Monte Carlo based property uncertainty analysis is performed, as described in 2.3, is performed, to analyze the real working fluids including their corresponding property uncertainties. The Monte Carlo results provide a cumulative distribution function for the model output (i.e. COP) of each of the real fluids. And the 95%-confidence interval of the COP output with respect to the corresponding property parameter values can be obtained for every real fluid.

Chapter 4

Results and discussion of property uncertainty analysis

The results of the the property uncertainty analysis applications are discussed here. Firstly, the outcomes from the study on the development of group contribution based property models for safety-related properties are shown. Secondly, the results from the uncertainty assessment of SRK, Peng-Robinson and PC-SAFT EoS are compared.

4.1 New prediction methods for safety-related properties

Here the main results of the application of the methodology for the development and uncertainty analysis of GC based property models are highlighted. The details are given in **Papers (A)**, **(B)** and **(C)** in Part II.

In the first part insights concerning the analysis of the described methodology itself are presented. In the second part the developed GC methods for safety-related properties are considered.

4.1.1 Analysis of methodology for development and uncertainty analysis of GC methods

Regression models

The performance of the applied regression models for the heat of combustion, ΔH_c°, GC method is shown in Table 4.1. The results are depicted *before* and *after* outlier deletion, where N_{out} is the number of outliers removed. R^2 is the Pearson correlation coefficient and *ARE* represents the average relative error between the experimental and the predicted data. Furthermore, P_{rc} gives the percentage of the experimental data points found within $\pm$25% relative error range of the model [16].

Table 4.1: Regression model performance statistics, the best value of the respective column is highlighted.

	R^2	N_{out}	*ARE*	P_{rc}
Ordinary least-squares *before* outlier deletion	0.99	0	1.10	99.75%
Robust regression *before* outlier deletion	0.99	0	0.75	99.62%
Weighted least squares *before* outlier deletion	0.99	0	1.82	99.62%
Ordinary least-squares *after* outlier deletion	0.99	0	0.52	**100**%
Robust regression *after* outlier deletion	0.99	0	**0.50**	99.87%
Weighted least squares *after* outlier deletion	0.99	0	0.57	**100**%

As it can be seen in 4.1, the outlier deletion has improved the regression performance for ΔH_c°. After the outlier deletion, the results are relatively close for the

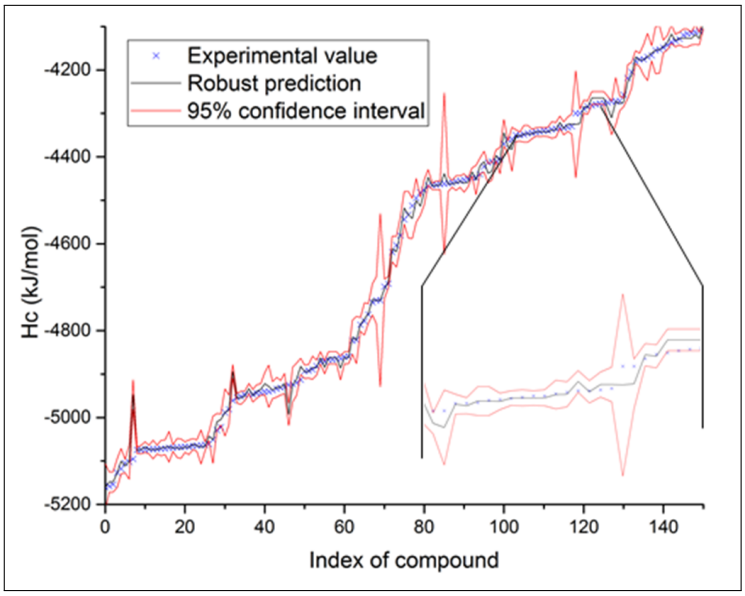


Figure 4.1: Experimental as well as predicted value of ΔH_c° for every compound with 95%-confidence intervals generated by covariance-based uncertainty analysis (robust regression without outliers). A section of the plot is enlarged to show the distribution of the experimental values around the prediction.

three models. The best fit according to the *ARE* is achieved by robust regression after outlier deletion.

Uncertainty analysis

Figure 4.1 shows the experimental and the predicted values of ΔH_c° with the respective 95%-confidence interval of the prediction for every substance for the covariance-based uncertainty analysis. As an example the prediction based on parameter estimates obtained using robust regression is shown. The compounds are ordered from lowest to highest value and given an index number respectively. The confidence intervals are individual for each compound. The trend in Figure 4.1 is a narrow band along with the experimental values. As it can be seen the experimental values lie within the calculated 95% confidence intervals.

Parameter identifiability analysis

The consideration of the 95%-confidence interval of the parameter estimates, allows evaluating the practical identifiability of the GC factors. There is a large

number of parameters that had a large confidence interval corresponding to a relative parameter estimation error being larger than 50%. For example for the use of robust regression for ΔH_c° 95 out of 235 parameters failed practical identifiability. The practical identifiability depends on two main issues: The amount of data for the parameter estimation and the correlation between parameters.

If there are sufficient data points to calculate the parameter estimates, the confidence interval gets smaller and hence, the parameters are practically identifiable. However, in GC parameter estimation there might be several functional groups that only occur in very few compounds.

The second major source of parameter identifiability problems is high correlation (>0.7) between parameters. In GC methods, correlation is intrinsically often the case, because certain functional groups can occur frequently together (depending on the data set) [90].

In many property modelling studies, practical identifiability of parameters has either not been considered or neglected. The first implication of this is that the estimated parameter values should not be attributed physical meaning, since their values are not unique. Second, for practical application purposes, it is desirable to keep the parameters in the model, because in this way the application range of the GC model is higher.

Whenever a model with poorly identifiable parameters is used, the uncertainty of the prediction becomes crucial. The confidence interval of property prediction provides a measure of the prediction quality of the model developed, which the end user can use to judge if the prediction uncertainty is suitable for the intended application.

Effect of addition of higher order groups on property value and uncertainty

It is valuable to analyze, what the influence is of the correlated parameters on the prediction and on its uncertainty. The results obtained in this study showed that high correlation influences the mean prediction but not the uncertainty bounds (the upper and lower 95% confidence interval).

In 155 out of 794 molecules the introduction of 2nd or 3rd order groups increases the relative error between experimental and predictive values for more than 10%. This particularity is exemplified and investigated by the compound cis,trans-2,4-hexadiene. The parameter correlation matrix given in Table 4.2 shows that the GC factors of cis,trans-2,4-hexadiene are highly correlated. The prediction and 95%-confidence interval for the selected substance (shown in Table 4.3) considers 1st order only, 1st and 2nd order as well as 1st, 2nd and 3rd order GC factors. While adding more groups, the relative error of prediction for cis,trans-2,4-

Hexadiene compound increases. However, it does not affect the calculation of the 95%-confidence interval of the property prediction.

Table 4.2: Parameter correlation matrices for cis,trans-2,4-Hexadiene. The red color indicates a positive correlation of higher than 0.7 and the orange color indicates a negative correlation lower than -0.7.

GC-factors	ΔH_c°	'CH3'	'CH=CH'	'CHn=CHm-CHp=CHk'	'CH3-CHm=CHn'
ΔH_c°	1.00				
'CH3'	-0.96	1			
'CH=CH'	-0.02	0.03	1		
'CHn=CHm-CHp=CHk'	0.02	-0.03	-0.94	1.00	
'CH3'-CHm=CHn'	0.02	-0.07	-0.96	0.86	1.00

Table 4.3: Prediction and 95%-confidence interval for a selection of substances comparing the usage of only 1st order GC-factors with the usage of 1st and 2nd as well as 1st, 2nd 3rd order groups.

	Relative error between prediction and experimental value			Boundary of 95%-confidence interval		
Used GC-factors	1st	1st, 2nd	1st, 2nd, 3rd	1st	1st, 2nd	1st, 2nd, 3rd
cis,trans-2,4-Hexadiene	0.024	0.038	0.029	±13.96	±13.93	±13.14

This particularity can be understood by looking at Eq. 2.3 and 2.4 (see section 2.1). The first reason lies in the negative correlation. If two parameters are negatively correlated and have similar sensitivity to the model output (corresponding to the Jacobian), their uncertainties will tend to cancel eachother out [90]. The second cause is the nature of the calculation of mean sum of squared error $S = SEE/(n-p)$, because the relative decrease of SSE is compensated by the corresponding increase in the number of parameters used for its estimation.

As a result, one can conclude that definition and inclusion of higher groups for a GC model may not always lead to a more accurate property prediction. However, the 95%-confidence interval does not enlarge due to poor parameter identifiability. It is therefore highly recommended that developers and users of GC models in general always state the 95%-confidence interval.

4.1.2 Comparison of the new GC models with other property estimation models

The average relative error *ARE* and the number of data included for all the newly developed GC methods are compared to already existing property prediction models. The details of this comparison are shown in **Paper (A)** for ΔH_c° and in **Paper (B)** in Part II for *LFL*, *UFL*, $LFL(T)$, *FP*, *AIT*. Here, the comparison of the newly developed *LFL* GC method is shown in Table 4.4.

The comparison could only be made according to the average relative error *ARE*, due to the fact that no uncertainty analysis has been performed by the other authors. The current model provides for every predicted value the corresponding uncertainty, which is lacking in all the other models.

Table 4.4: Comparison of developed LFL model with existing GC models. Abbreviations: average relative error (*ARE*), Marrero/Gani (MG), group contribution (GC), atom and bond connectivity (AC), quantitative structure property relationship (QSPR), artificial neural networks (ANN).

	Model structure	*ARE* [%]	No. of data
New method	**MG GC**	**12**	**443**
Oehley, 1953 [91]	AC	27	-
Solovev et al., 1960 [92]	GC	25	-
Shimy, 1970 [93]	CN	24	9
Shebeko, 1970 [94]	AC	21	70
Seaton, 1991 [95]	GC	16	152
Kondo et al., 2001 [96]	GC	24	238
Albahri, 2003 [97]	struct. GC	10	109
Gharagheizi, 2008 [98]	QSPR	8	1056
Pan et al., 2009 [99]	QSPR	5	1038
Gharagheizi, 2009 [100]	ANN	4	1056
Lazzús, 2011 [24] [101]	ANN	9	328
Rowley et al., 2011 [28]	GC	11	509
Bagheri et al., 2012 [102]	QSPR	1	1615
Mendiburu et al., 2015 [103]	semi-empirical	9	120

Considering the *ARE* of *LFL*, the model developed in this study performs better than the majority of the existing *LFL* models and took also more data into account for parameter estimation (i.e. a wide application range can be achieved). The current *LFL* model performs similar in comparison to the recent GC prediction method of Rowley et al. [28] and the best performing model of Albahri [97]. The work of Mendiburu et al. [103] took only C-H compounds into account. The

ANN (artificial neural network) methods of Lazzús [101] and Albahri [97] show better performance statistics as well. However, these authors took a lower amount of experimental data points into account for the fitting of their model (giving a narrow application range). Furthermore, an ANN structure is very complex for even a relatively small number of fitted data.

The ANN and QSPR (quantitative structure property relationship) models of Gharagheizi [100, 98], Pan et al. [99] and Bagheri et al. [102] for LFL have a much lower ARE and more data points. However, the amount of data consists of all experimental data and predicted values available in the DIPPR database [31], which is not a scientifically accepted way to compare model performance statistics. A parameter estimation should only be based on experimental data points only [104]. Furthermore, GC models allow adding new experimental values to the parameter estimation without changing the model structure. In QSPR and ANN, model building needs to be performed all over again.

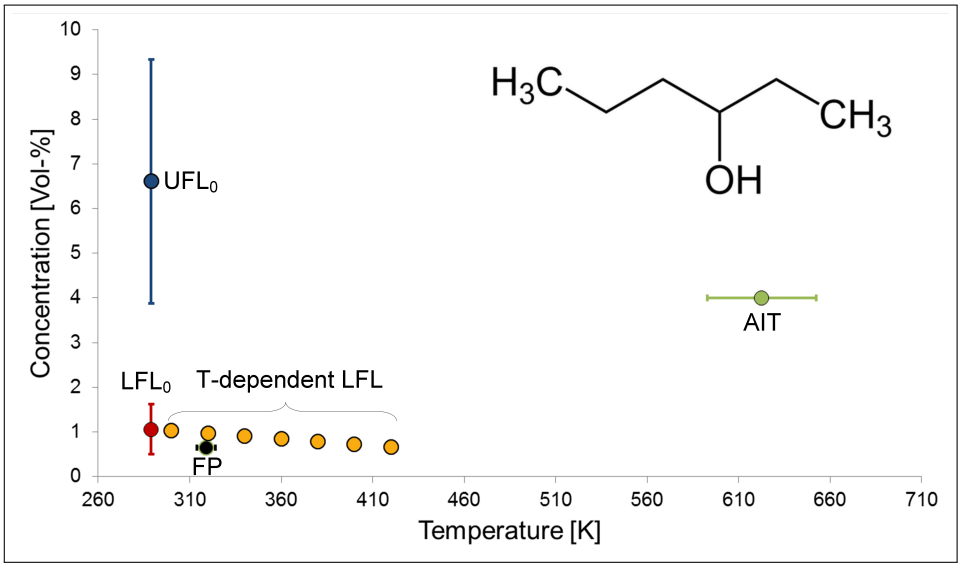


Figure 4.2: Overview of the generated flammability-related properties by the developed GC MG models including 95% confidence interval: LFL_0 (lower flammability limit at T=298K), UFL_0 (upper flammability limit T=298K), FP (flash point), AIT (auto ignition temperature) and temperature-dependent LFL (without uncertainty).

4.1.3 Demonstration of model application

The developed models allow calculating the safety properties from the molecular structure only and include an uncertainty analysis. Figure 4.2 depicts the result

of example calculations with the developed GC models for 3-Hexanol. As it can be seen the newly developed GC methods provide an overall picture of the major flammability property predictions including the corresponding 95% confidence interval.

The details of the example calculations for 3-Hexanol as well as predicted values for a variety of organic compounds are provided in **Paper (B)**. The GC models form the basis for a new property prediction software tool called SAFEPROPS which is able to estimate major safety-related and environmental properties for organic compounds [105].

4.2 Comparison of SRK, Peng-Robinson and PC-SAFT equation of state uncertainty propagation

An overview of the results of the assessment of the equations of state (EoS) with respect to the corresponding EoS parameter uncertainty is given here. A detailed analysis is provided in **Papers (D)** and **(E)** in Part II.

4.2.1 Overview of the output uncertainties in log(P)-h and T-s diagrams

The outcome of the Monte Carlo methods is shown on temperature-entropy (T-s) and logarithmic pressure-enthalpy (log(P)-h) diagrams in Figure 4.3, which gives an overview of all the uncertainty analyses for the SRK EoS (left hand side) and the PC-SAFT EoS (right hand side). The uncertainty is a varying band for both the saturation curves (yellow) and the cycle design (red). The larger the width of the band, the greater the uncertainty.

From the overview figures, it is possible to visually analyze the results of the fluid-specific EoS parameter uncertainty propagation. For example, from the output uncertainty from the SRK EoS shown on the T-s diagram, the expansion process uncertainty (states 5 to 6) is larger than the uncertainty in the evaporation line (states 4 to 5). For the PC-SAFT EoS, a comparatively wide band can be seen for the evaporation temperature (states 2 to 5) as well as for the saturated liquid line (states 3 to 4) on the T-s diagram. However, note that the uncertainties of PC-SAFT and SRK cannot be compared directly using Figure 4.3 because the outputs are normalized by the different EoS mass flow rates.

4.2.2 Ideal-gas contribution versus departure function

The effects of the parameter uncertainties on the ideal-gas contribution (i.e., the heat capacity expression) can be compared to those from the departure functions (i.e., SRK, Peng-Robinson and PC-SAFT EoS). For the comparison SRK and Peng-Robinson are considered in their common parametrization (T_c, T_c and ω). Figure 4.4 shows the output distributions of the ORC net power output W_{NET} as obtained from the evaluated Monte Carlo samples.

The results of the combined uncertainty propagations of the departure functions (SRK and PC-SAFT) and the ideal-gas contributions are shown together with the results from the uncertainty analysis, when only the departure functions or the ideal-gas contributions are varied subject to their uncertainties. In Figure 4.4 the

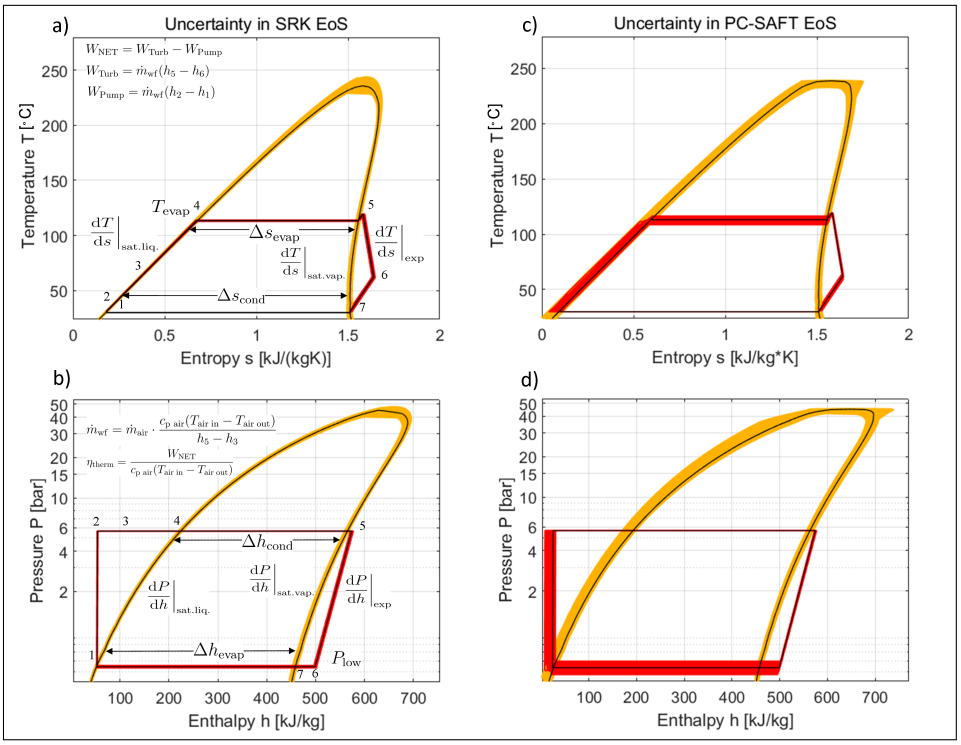


Figure 4.3: Representation of uncertainty with respect to the fluid properties in the T-s diagram and log(P)-h diagram for cyclopentane for SRK and PC-SAFT input uncertainty: Monte Carlo simulations overlaid (yellow/red) and mean (solid black line). The numbers refer to Figure 3.1. In addition to the ORC model output variables described in chapter 3.1.2, the following outputs are considered in detail in (Paper (D): the evaporation temperature (T_{evap}), the condensation/evaporation entropy as well as enthalpy (Δs_{cond}, Δh_{cond}, Δs_{evap}, Δh_{evap}, as well as the various slopes for the T-s and log(P)-h) diagrams (e.g. the slope of the expansion line in log(P)-h diagram $dP/dh|_{exp}$).

propagated input uncertainties of the three departure functions are compared to the ideal-gas contribution.

Considering the differences in the widths of the distributions of the net power output W_{NET} in Figure 4.4, the influence of the propagated heat capacity uncertainties on the model output is small compared to the effect of the uncertainties in the departure functions for SRK, Peng-Robinson and PC-SAFT. Hence, the sensitivity of the departure functions is much larger than that of the ideal-gas contribution for all the output variables. The same conclusion is found for the other output variables (see **Paper (D)**).

The result is expected, since in the ORC both gas and liquid states exist at high

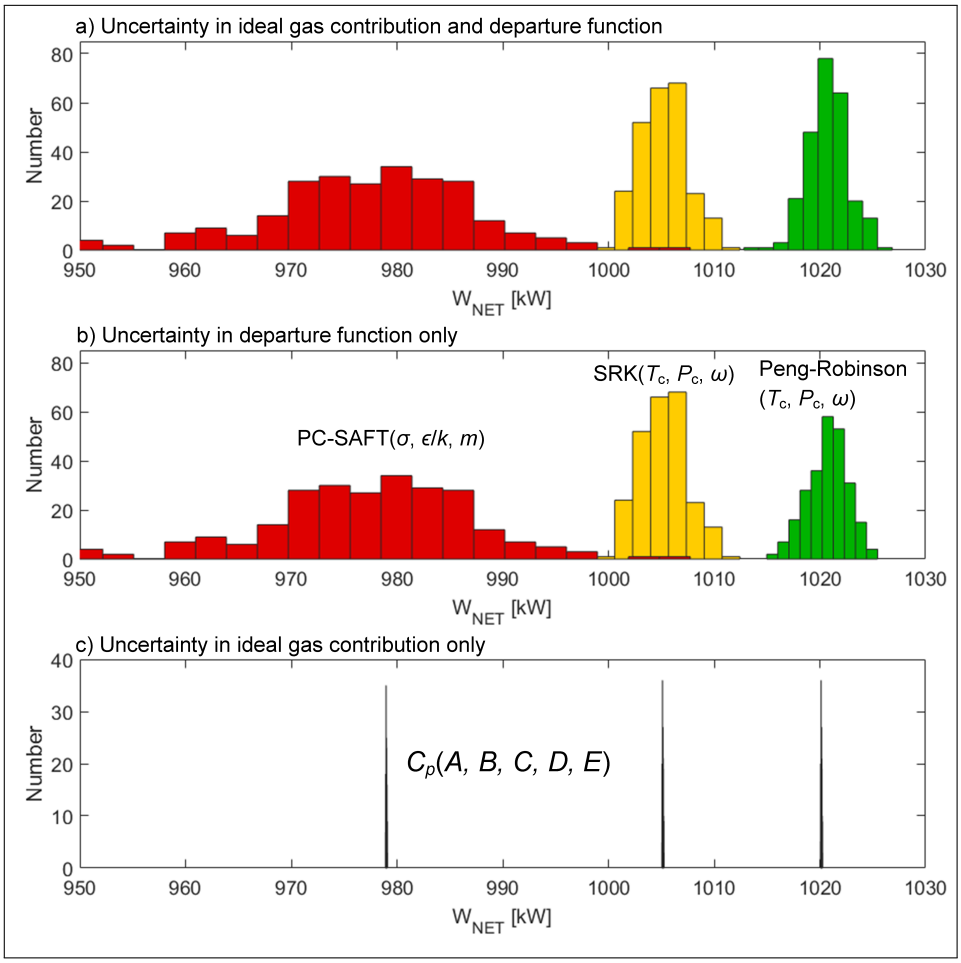


Figure 4.4: Output distributions of the ORC net power output W_{NET} from Monte Carlo simulations. Subfigures a, b and c compare the output distributions of the propagated input uncertainties of the departure functions PC-SAFT (red), SRK (yellow) and Peng-Robinson with the ideal-gas contribution (i.e. from heat capacity parameter uncertainties).

pressures. Therefore, the real-gas deviation from the ideal-gas becomes important.

4.2.3 Cubic versus PC-SAFT equation of state

Figure 4.5 shows the output distributions of the ORC net power output W_{NET} as obtained from the evaluated Monte Carlo samples. The results of the combined uncertainty propagations of the departure functions of SRK, Peng-Robinson and PC-SAFT EoS are shown in three parts: The upper subfigure shows the distribution of PC-SAFT parameterized in σ, ε/k, and m, along with the cubic EoS (SRK and Peng-

Robinson) parameterized in T_c, P_c and ω. The middle and the bottom subfigures depict the distributions of SRK and of Peng-Robinson parameterized in a, b and β.

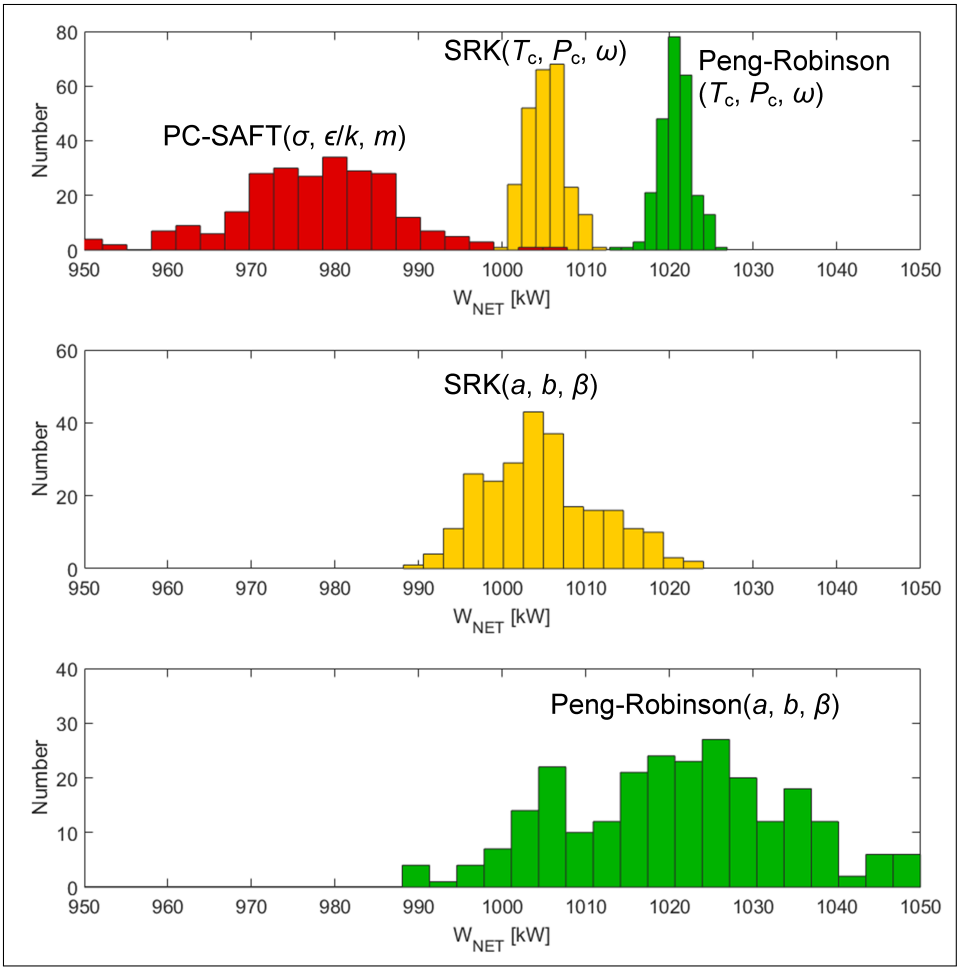


Figure 4.5: Output distributions of the ORC net power output W_{NET} from Monte Carlo simulations. The subfigures compare the output distributions of the propagated input uncertainties of the departure functions SRK (yellow), PC-SAFT (red) and Peng-Robinson (green). Distributions of SRK and Peng-Robinson EoS are shown when parameterized in both (T_c, P_c and ω) and (a, b and β).

Figure 4.5 reveals several aspects. The common parametrizations of the SRK and Peng-Robinson EoS models (T_c, P_c ω) yield in a narrower uncertainty distributions than PC-SAFT. This indicates that, from an uncertainty point of view, the SRK and Peng-Robinson EoS could be preferable for the given application, i.e. for the performance evaluation of a working fluid in ORC process design. The cause of this could be the differences in mathematical form. The PC-SAFT parame-

ters enter into several different functions, which are (from a model point of view) highly nested and often of contrasting effects, whereas the common SRK and Peng-Robinson parametrization is more direct.

However, the uncertainties from SRK and Peng-Robinson (parametrized in T_c, P_c ω) have not been obtained from the same data as PC-SAFT. When the cubic EoS models are reformulated with three regressed parameters (a, b, β), and fitted to the same data as the PC-SAFT EoS, the uncertainty distributions became much broader, though the mean values where quite similar.

This differences according to the parametrizations of the cubic EoS seems to be due to the greater amount of data used in the regression for the (a, b, β)-forms. Even though the uncertainty distributions are different, the mean values for the cubic EoS are essentially the same for both parameterizations. The later insights suggest that the uncertainty propagation from properties to mean process outcomes is determined more by the model form than by data used in regressions.

Second, the EoS models lead to different mean values for W_{NET} with the PC-SAFT giving the lowest by 5%. There is significant overlap of the distributions only for the (a, b, β) forms of the cubic EoS. Thus, the process output results do depend strongly on the EoS model form. In a different study Leekumjorn et al. [106] also concluded that the relative errors for both PC-SAFT and cubic EoS showed deviations of 2-6% for a variety of hydrocarbon fluids.

In this work it could also be shown that the uncertainties in W_{NET} are significant (of the order of 2 – 5%), and need to be recognized when designing thermodynamic cycles.

It is suggested that future process modelling studies should examine uncertainty as well as accuracy of potential EoS models in order to gain additional insights about uncertainties in fluid properties, parameters, and EoS model structures. In particular, measurement errors in data should be taken into account when developing and reporting EoS models and the resulting covariance matrix of model parameters should be calculated and reported.

This allows direct propagation of parameter uncertainties to model output uncertainties, which provides another and important criterion for property model selection for process design.

Chapter 5

Results and discussion of working fluid selection

This chapter shows the results of the two case studies on fluid selection and design under property uncertainties. In the first part, the outcomes of the case study on fluid selection under property uncertainties for an ORC system are presented. In the second part, the results of the novel reverse engineering approach for fluid design based on sampling and uncertainty analysis are considered.

5.1 Fluid selection under property uncertainty for organic Rankine cycle for low-temperature heat recovery

5.1.1 Ranking of working fluids

The results of the case study for working fluid selection under property uncertainty for an organic Rankine cycle (ORC) are presented here (see also section 3.2.1). After screening all of the compounds of the DIPPR 801 database, an uncertainty analysis with respect to the respective properties has been performed for all feasible compounds. Having obtained the mean and the 95%-confidence interval of the net power output of the fluids, it is possible to rank the compounds.

Figure 5.1 shows the mean value of the distribution of the ORC net power output W_{NET} for the 20 best performing compounds for the given ORC power plant. The table includes the corresponding uncertainty (95%-confidence interval) with respect to the property inputs. The ranking does not include safety and environmental properties of the fluid, because the particular focus of this study lies in the analysis of property uncertainty. However, it has to be acknowledged in the current case that the fluoro-compounds have a relatively high Global Warming Potential ($GWP>$ 150) and that the hydrocarbons have a high lower flammability limit ($LFL>0.1$ kg/m^3).

Alternatively the fluids can be ranked according to their respective lower bound of the 95%-confidence interval (see Figure 5.2). This is a conservative approach of ranking and can be considered as the statistically robust way to identify promising working fluid candidates.

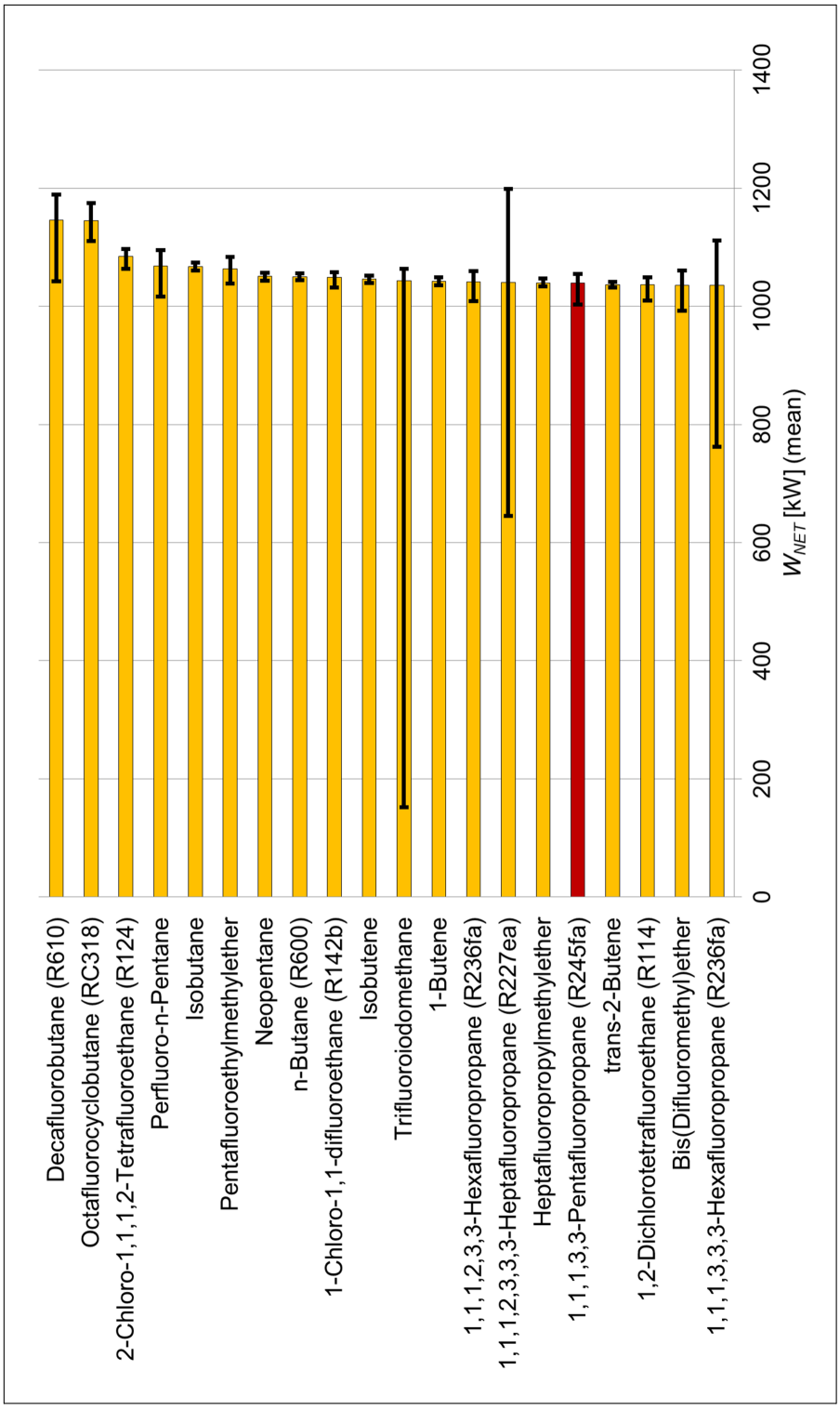


Figure 5.1: Mean value of the net power output W_{NET} of the 20 best performing compounds including the 95%-confidence interval (thin black bars) obtained from the uncertainty analysis with respect to the fluid properties. The red bar corresponds to the base case compound.

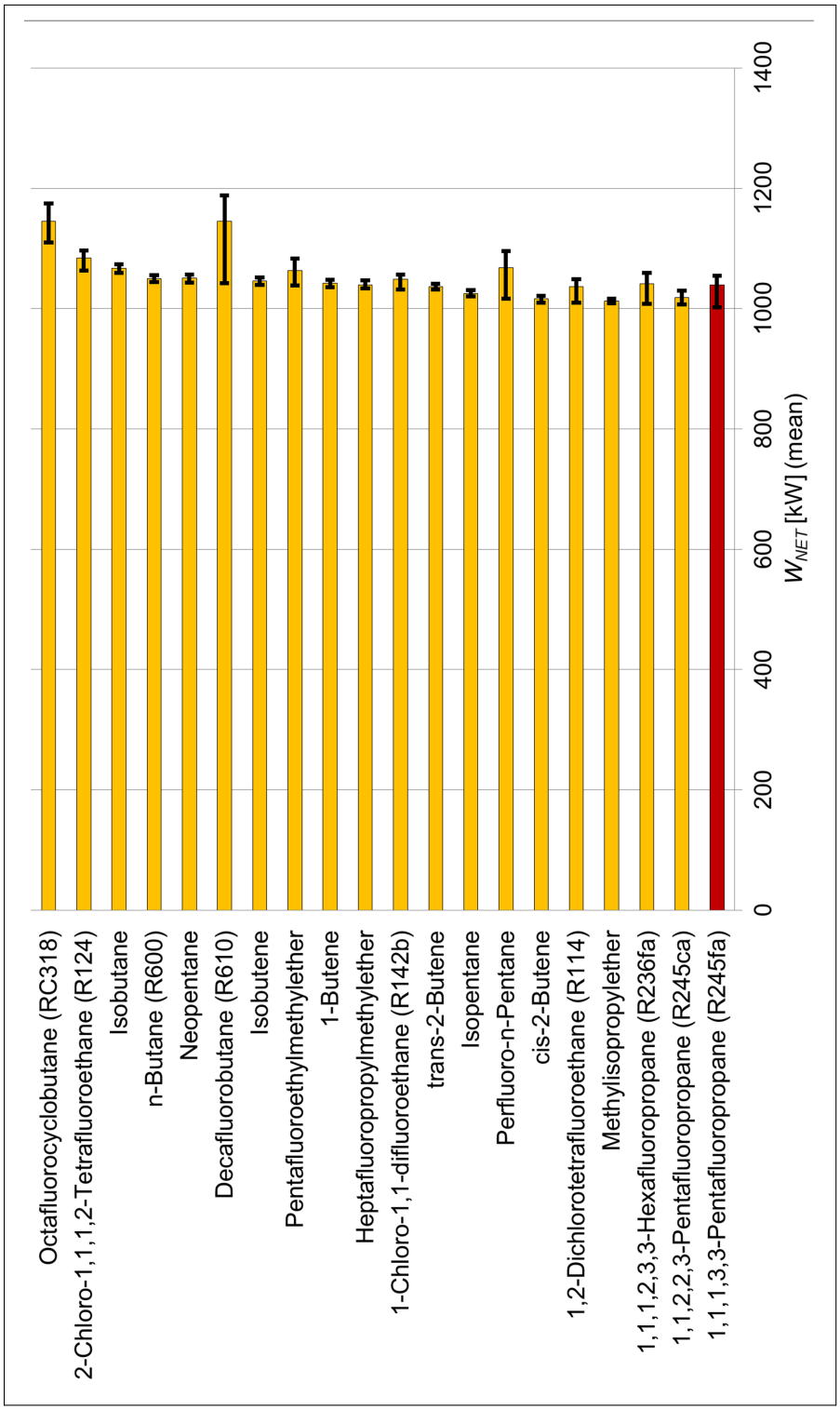


Figure 5.2: Ranking according to the lower bound value of the 95%-confidence interval of the net power output W_{NET}. The red bar corresponds to the base case compound.

5.1.2 Fluid selection under uncertainty

Knowing the 95%-confidence interval of the net power output for the screened chemical compounds gives an important new dimension in the preliminary selection of suitable working fluid candidates. Some working fluids, e.g. isobutane (5th compound from top in Figure 5.1), have a very small model output uncertainty range, whereas others, e.g. trifluoroiodomethane (11th compound), have a very large.

If the 95%-confidence interval of a compound overlaps with the one of the base case, it is statistically impossible to say, which of them performs better. This is the case for one of the two top compounds. The mean value of decafluorobutane (1st compound in Figure 5.1) and octafluorocyclobutane (2nd compound) are very close to each other, but the 95%-confidence interval of decafluorobutane is overlapping with the 95%-confidence interval of the base case (see Figure 5.1). This can also be seen in Figure 5.2, where the ranking has been made according to the lower bound of the 95%-confidence interval, as decafluorobutane is not anymore at the top.

Hence, the uncertainty analysis provides important additional information for the interpretation of the results. Based on the analysis of this study, the best performing compound with the smallest uncertainty range is in fact octafluorocyclobutane. Hence, this fluid can be selected for the given cycle. However, the study also implies that more reliable property data for decafluorobutane is needed.

There are two major causes for large net power output uncertainty:

1. The input property uncertainty of one or more parameters is high and results in a large net power output uncertainty. This is directly related to the reliability of the measured and predicted property data.

2. The cycle is operated in a sensitive region in terms of the fluid properties with respect to the model evaluation for a particular fluid. Hence, small variations of the parameters have a large impact on the model output. The knowledge of whether a fluid is sensitive to the ORC model structure or not is a priori unknown. Therefore, an uncertainty analysis of the model output with respect to the fluid properties can give vital information.

The range of uncertainty can be considered as a novel criterion in model based working fluid selection. The narrower the 95%-confidence interval, the more reliable the property data and the less sensitive the fluid performs in the cycle. This information can be vital for further detailed modelling and experimental validation studies of identified promising fluids.

5.2 Reverse engineering approach for working fluid design of heat pump system for low-temperature heat recovery in food industry

5.2.1 Ranking of identified cyclic hydrocarbon working fluids

The new reverse engineering approach based on Monte Carlo sampling has been applied for an industrial heat pump system used for waste heat recovery from spray drying facilities in dairy industry, focusing in particular on potential novel cyclic hydrocarbon working fluids. Here the main outcomes of the study are presented, the detailed results can be found in **Paper (G)**.

Table 5.1 shows the coefficient of performance (*COP*) mean value of the distribution for the best performing compounds for the considered heat pump cycle. It includes the uncertainty with respect to the property input, which has been propagated through the cycle. Additionally, the *COP* of the closest top performing virtual sample fluid is shown. The *COP* mean value and the corresponding uncertainty ranges (95%-confidence interval) are also represented in Figure 5.3. In order to compare the identified cyclic hydrocarbons, the *COP* has also been calculated for 3 commonly used refrigerants that would be suitable for the given process: R-152, R-143 and R-245fa. The ranking includes the lower flammability limit obtained from the prediction method developed and presented in section 3.1.1. The global warming potential (*GWP*), calculated by the method of Hukkerikar et al. [21], is very small for all of the considered cyclic compounds ($GWP<0.1$) compared to commonly used fluoro-hydrocarbon refrigerants, as e.g. R-152 ($GWP=53$), R-143 ($GWP=353$), R-245fa ($GWP=1030$) [107].

According to the recently published F-gas regulation of the European Union working fluids with a GWP of higher than 150 should be phased out [12]. Furthermore, the ozone depletion potential of all of this cyclic hydrocarbon working fluids is zero [21].

The molecules identified by the algorithm are comparatively small cyclic hydrocarbons. Large cyclic, aromatics or polycyclic compounds had in particular a critical temperature, which was much larger than the critical temperature of the optimal virtual fluids.

The *COP* uncertainty range of the identified compounds is overlapping with the 95%-confidence interval. The best performing compound with the smallest uncertainty range in *COP*, a low *GWP* (<150) and a comparatively lower flammability (for safer operation) for the considered cycle is cyclopentane.

Table 5.1: Best performing compounds ranked by *COP* mean value including uncertainty.

Rank	Working fluid name	Mean *COP*	*COP* 95%-confindence interval		*COP* of closest optimal virtual fluid	LFL [Vol-%] (with ASHRAE Safety group [108])
1	Cyclopentane	3.06	3.00	3.11	3.17	1.41±0.90 (A3)
2	Cyclobutane	3.04	2.95	3.12	3.17	1.41±0.59 (A3)
3	cis-1,2-Dimethylcyclo-propane	3.04	2.87	3.11	3.17	1.40±0.58 (A3)
4	Methylcyclo-butane	3.03	2.86	3.11	3.17	1.78±0.90 (A3)
5	trans-1,2-Dimethylcyclo-propane	3.00	2.84	3.09	3.17	1.31±0.57 (A3)
6	Methylcyclo-propane	2.97	2.88	3.05	3.17	1.78±0.50 (A3)
7	1,1-Dimethyl-cyclopropane	2.86	2.63	3.08	3.19	0.69±0.59 (A3)
8	Butylcyclobutane	2.73	2.38	3.03	3.16	0.76±0.90 (A3)
9	1,1-cis-3,4-Tetramethylcyclo-pentane	2.49	2.25	3.01	3.18	0.51±0.60 (A3)
10	1,1,2-Trimethyl-2-ethylcyclopropane	2.48	2.14	2.93	3.16	1.40±0.59 (A1)
	R-152 (1,2-Difluoroethane)	3.08	3.03	3.11	-	4.15±0.41 (A1)
	R-143 (1,1,2-Trifluoroethane)	3.07	3.00	3.12	-	6.20±1.55 (A1)
	R-245fa (1,1,1,3,3-Pentafluoro-propane)	3.08	3.05	3.09	-	7.70±0.77 (A1)

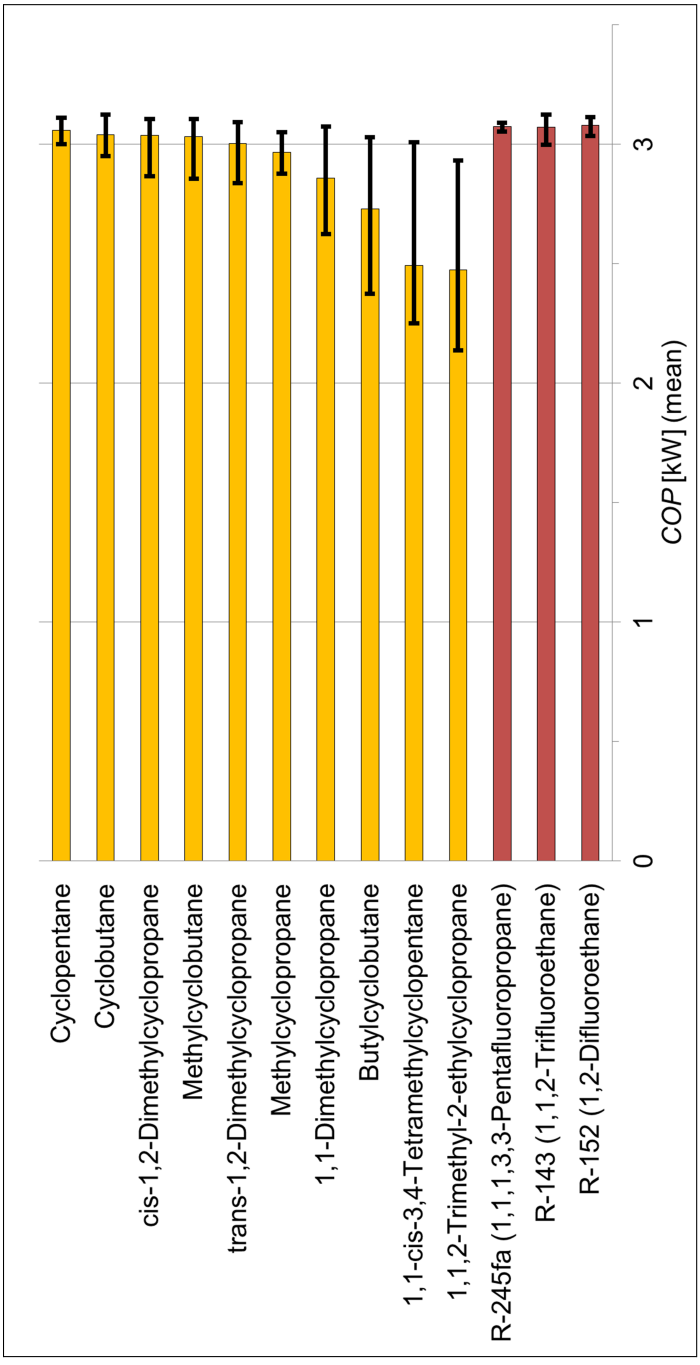


Figure 5.3: Ranking according to mean value of the *COP* of the best performing compounds. The 95%-confidence interval (thin black bars) was obtained from the uncertainty analysis with respect to the fluid properties, the identified cyclic hydrocarbons (yellow) are compared to commonly used refrigerants (red).

The identified compounds have similar *COP* as commonly used refrigerants. However, cyclic compounds have a much lower impact on the climate in agreement to present regulations on working fluids. However, due to their high flammability, safety-measures need to be considered when the compounds are further investigated experimentally. In particular small rings (e.g. cyclopropane- and cyclobutane-compounds) often suffer from ring tensions and can be instable [109], which is also reflected in their small value for the lower flammability limit.

5.2.2 Discussion of reverse engineering approach

Considering the upper bound of the 95%-confidence interval of the real fluids in Figure 5.3, it can be seen that the *COP* values of the identified working fluids come close to the *COP* value of the optimal virtual fluids. This implies that the sampling based reverse engineering approach succeeded in identifying real working fluids, which could provide a high *COP*.

This identification has been achieved without evaluating all considered 2126 cyclic compounds in the cycle model and without solving a product design optimization problem. This adds credit to the effectiveness of the Monte Carlo sampling concept employed in the novel methodology for screening large chemical database. Furthermore, property uncertainty information is taken into account for the analysis of the performance of the identified fluids.

It is observed that cyclic hydrocarbons, which have had a distance function $d_{xy}^{\text{tot}} > 0.2$, have been too far away from the virtual fluids to give a feasible solution. This boundary corresponds to an average relative difference between the property value of the virtual fluid and the real fluid of 20%. This value is specific for the given study. However, it confirms that the property search space is highly non-continuous, which makes standard (non-sampling based) numerical optimization tools [110] difficult or even impossible.

The upper bounds of the identified real fluids are very close to each other, whereas the lower bounds differ largely. This can be explained by the algorithm for the calculation of the distance value between the property of the real and the virtual fluid. The distance is the difference between the property value of the virtual fluid and the boundary of the uncertainty range of the real fluid. Hence, a statistically optimistic approach is taken for the identification of the working fluids. This optimistic approach is reflected in the uncertainty range of the model output (*COP*), where the upper bound shows the statistically best possible performance of the real fluid, which is based on the working fluid properties closest to the optimal virtual fluid. However, the uncertainty analysis after the reverse engineering algorithm considers the property uncertainty over the whole range from lower to upper

bounds. In this way, it is revealed which fluids can (statistically) also have a lower performance (conservative approach). Hence, the property based uncertainty analysis is an important complementary information after the identification of suitable working fluids using the sampling based reverse engineering approach.

The properties of the optimal virtual fluids can be found in **Paper (G)** in Part 2. The identified properties of the best performing virtual fluids can also serve as target properties for further studies, e.g. the identification of optimal mixture compositions. Hence, the sampling based approach for reverse engineering is not limited to the example of pure component fluid design shown in this work.

The study gives reason to believe that the novel reverse engineering approach can be useful for process developers of thermodynamic cycles, because it allows a simple identification of working fluids through the application of Monte Carlo methods. Furthermore, the methodology has been formulated in a generic way and it is possible to apply it to product-process design problems in process engineering beyond working fluids for thermodynamic cycles.

Chapter 6

Conclusion and perspectives

6.1 Main conclusions

This thesis project developed a comprehensive framework with methodologies and tools to quantify and propagate working fluid property uncertainties on every level of a thermodynamic cycle model. Furthermore, new methods for the identification of pure component working fluids under property uncertainties were suggested and applied to industrial case studies.

The systematic methodology for parameter estimation and uncertainty analysis of group contribution (GC) based property models was applied to the development of novel prediction methods for flammability-related working fluid properties. The developed flammability limit models have a higher accuracy than existing GC models and are much simpler to apply than current neural network or quantitative-structure-property-relation (QSPR) models.

The analysis of the GC framework revealed that the advanced parameter estimation using robust regression and the systematic outlier treatment using the empirical cumulative distribution function (CDF) provided an improved performance statistics compared to classical non-linear regression. The detailed investigation also showed that GC-based models can suffer from parameter identifiability issues characterized by significant correlation between estimated parameters and a large confidence interval. However, it was also demonstrated that the GC-based models still can be used successfully, when the 95% confidence interval of the model predictions are also calculated and reported.

It was also shown how parameter uncertainties for two types of equations of state (EoS), cubic and PC-SAFT, could be obtained from measured data using a bootstrap method. Furthermore, a Monte Carlo procedure propagated the uncertainties onto the process model output of an organic Rankine cycle (ORC). In the

case study it was found that the propagated output uncertainties of the ORC model were determined more by uncertainties in the EoS departure functions than uncertainties from the ideal-gas contribution from the heat capacity model. Comparing the uncertainties of the departure functions revealed that common parametrizations of the cubic EoS models gave narrower uncertainty distributions than PC-SAFT. However, when the two cubic EoS were reformulated with three regressed parameters, and fitted to the same data as the PC-SAFT, the uncertainty distributions became much broader. Hence, the model structure of an EoS is crucial for its uncertainty propagation.

The case study on fluid screening and selection under property uncertainties for a low-temperature ORC revealed that ranking of working fluids could be significantly different based on whether or not the mean value or the uncertainties (e.g. the lower bound of the 95%-confidence interval) of the model output were considered. Hence, uncertainty analysis with respect to the input property uncertainties is a vital tool for ORC model analysis and working fluid selection.

The novel sampling based reverse-engineering method for working fluid design has been developed and applied to an industrial heat pump system used for heat recovery in food industries. The method was applied in order to identify cyclic hydrocarbon based working fluids. Cyclic hydrocarbons, which achieved a *COP* value around 3, were suggested. However, the calculation of the lower flammability limit also showed that the identified cyclic fluids are highly flammable. The new approach provides an alternative to classical optimization-based problem formulations and takes into account property uncertainties.

6.2 Perspectives

The work of this thesis showed that it is necessary that developers of property models, such as GC methods, state the prediction including estimation of its uncertainty (i.e. the 95% confidence interval).

In the domain of EoS, research is required on the uncertainty quantification and propagation of the widely used Helmholtz-based EoS. As the uncertainty analysis of three-parameter EoS, like cubic and PC-SAFT EoS, showed, the uncertainty propagation can vary depending on the type of EoS, the fluid and the application considered, which is important for users, who apply EoS in process simulation.

Uncertainty analysis for Helmholtz-based EoS may also reveal insights about parameter identifiability and can help developers to improve the reliability of the respective function. To this extent global sensitivity analysis methods that reveal the overall influence of a particular parameter on the EoS output could support the

EoS model building. Furthermore, the uncertainty quantification and propagation of EoS should become a criterion for selecting an EoS, along with accuracy (i.e. goodness-of-fit) and ease of usage.

In the domain of working fluids for thermodynamic cycles, the combination of optimization algorithms for computer-aided molecular design (CAMD), which construct molecules from a set of groups (of a GC method), and property uncertainties should be investigated. The optimization problem becomes complex, because every GC factor is associated with an uncertainty range. Howerver, a solution may be achieved by performing stochastic optimization.

Further research is also needed in order to investigate the suggested working fluids in the case study for the implementation in low-temperature heat recovery systems. Experimental measurements and detailed process simulations need to be carried out to test these fluids in real application.

The current work only considered pure component fluids. The methodologies shown in this thesis should be extended to the design of working fluid mixtures. Mixture properties are generally more difficult to predict, hence the uncertainty information, becomes crucial for the investigation of promising new compositions.

The analysis of uncertainties in process models is not restricted to the application of working fluids. The work of this thesis generally encourages that process developers in all domains of chemical and mechanical engineering take property uncertainties into account in their simulations. An example of where the methods and tools of this thesis may be useful, is the consideration of the influence of uncertainties in raw materials, when novel processes are developed in chemical industries. The output of alternative process configurations can vary subject to the consideration of property uncertainties. Hence, the quantification of the output uncertainty for different process alternatives can support decision-makers in process industries.

The outcome of this thesis clearly suggests that property uncertainties should be taken into account for process simulation applications both in academia and in industry, in order to support the model-based engineering for the decisions on process fluids and process design.

Chapter 7

Abbreviations and terms

Abbreviation	Definition
AIT	Auto Ignition Temperature [K]
ANN	Artificial neural network
ARE	Average relative error [%]
CDF	Cumulative distribution function
CAMD	Computer-aided molecular design
COV	Covariance matrix
Corr	Correlation matrix
EoS	Equation of state
FP	Flash Point [K]
GC	Group contribution
LFL	Lower Flammability Limit [Vol-%]
GWP	Global warming potential
ORC	Organic Rankine cycle
PC-SAFT	Perturbed-chain statistical associating fluid theory
QSPR	Quantitative-structure-property-relation
SD	Standard deviation
SSE	Sum of squared errors
SRK	Soave-Redlich-Kwong equation of state
UFL	Upper Flammability Limit [Vol-%]
Var	Variance

Symbol	Definition
a	Parameter of cubic EoS
b	Parameter of cubic EoS
c	Parameter of cubic EoS
C_j	First order Marrero/Gani group contribution factor
c_p	Ideal gas heat capacity [kJ/(molK)]
D_k	Second order Marrero/Gani group contribution factor
d_{xy}	Distance function between virtual fluid x and real fluid y
E_l	Third order Marrero/Gani group contribution factor
$F[\theta]$	Genenal function with parameters
h	Specific enthalpy [kJ/kg]
ΔH_c°	Heat of combustion [kJ]
J	Jacobian
K^{LFL}	Proportionality constant for temperature-dependent lower flammability limit
N_{data}	Number of data points
N_j	Number of first order Marrero/Gani group contribution factors
m	Chain length parameter of PC-SAFT EoS [-]
m_{wf}	Mass flow of working fluid
M_k	Number of second order Marrero/Gani group contribution factors
O_l	Number of third order Marrero/Gani group contribution factors
P	Pressure [Pa], [kPa] or [atm]
$\dot{Q}$	Heat flux [kW]
R	Ideal gas constant [J/(molK)]
s	Entropy [kJ/(kgK)]
T	Temperature [K] or [°C]
t	Value of the student t-distribution
V_m	Molar Volumen [m^3]

Continued on next page

Continued from previous page

Symbol	Definition
W_{NET}	Net power output of an organic Rankine cycle [kW]
w	Weight factor
y	Property value (general)

Greak symbol	Definition
α	Percentile of student t-distribution
$\alpha(T)$	Function in cubic EoS
ε/k	Energy parameter of PC-SAFT EoS [J/K]
$\hat{\varepsilon}$	Residual error
η	Efficiency [-]
ω	Acentric factor [-]
σ	Segment diameter for PC-SAFT EoS [Å]
θ	Parameters (general)

Bibliography

[1] Bühler, F.; Nguyen, T.V.; Elmegaard, B. Energy and exergy analyses of the Danish industry sector. *Applied Energy*, 184, 1447–1459, 2016.

[2] Cignitti, S.; Frutiger, J.; Zühlsdorf, B.; Bühler, F.; Andreasen, J.G.; Müller, F.; Haglind, F.; Elmegaard, B.; Abildskov, J.; Sin, G.; Woodley, J.M. Forbedring af industrielle processers energieffektivitet. *Dansk Kemi*, 97, 10, 2016.

[3] Larsen, U.; Pierobon, L.; Haglind, F.; Gabrielii, C. Design and optimisation of organic Rankine cycles for waste heat recovery in marine applications using the principles of natural selection. *Energy*, 55, 803–812, 2013.

[4] Zühlsdorf, B.; Bühler, F.; Mancini, R.; Cignitti, S.; Brian Elmegaard. High Temperature Heat Pump Integration using Zeotropic Working Fluids for Spray Drying Facilities. *12th IEA Heat pump conference*, 2017.

[5] Bühler, F.; Fridolin, M.; Huang, B.; Andreasen, J.G.; Elmegaard, B. Mapping of low temperature heat sources in Denmark. *ECOS 2015: 28th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems*, 2015.

[6] MAN Diesel & Turbo. Waste Heat Recovery System (WHRS) for Reduction of Fuel Consumption, Emission and EEDI. Tech. rep., 2012.

[7] Invernizzi, C.M. The Organic Rankine Cycle. In *Closed Power Cycles, Thermodynamic fundamentals and Applications*, chap. 3, pp. 117–176. Springer Verlang, London, 2013.

[8] Bao, J.; Zhao, L. A review of working fluid and expander selections for organic Rankine cycle. *Renewable and Sustainable Energy Reviews*, 24, 325–342, 2013.

[9] Linke, P.; Papadopoulos, A.; Seferlis, P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. *Energies*, 8, 6, 4755–4801, 2015.

[10] Papadopoulos, A.I.; Stijepovic, M.; Linke, P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. *Applied Thermal Engineering*, 30, 6-7, 760–769, 2010.

[11] Chen, H.; Goswami, D.Y.; Stefanakos, E.K. A review of thermodynamic cycles and working fluids for the conversion of low-grade heat. *Renewable and Sustainable Energy Reviews*, 14, 9, 3059–3067, 2010.

[12] Regulation (EC) No 1005/2009 of the European Parliament and of the council. *Official Journal of the European Communities*, L 269, September 2000, 1–15, 2000.

[13] Hwang, Y.; Ohadi, M.; Radermacher, R. Natural refrigerants. *Mechanical Engineering*, 120, 96–99, 1998.

[14] Poling, B.E.; Prausnitz, J.M.; O'Connell, J.P. The Estimation of Physical Properties. In *The Properties of Gases and Liquids*, pp. 1–9. McGraw-Hill, New York, 2004.

[15] Dong, Q.; Chirico, R.D.; Yan, X.; Hong, X.; Frenkel, M. Uncertainty Reporting for Experimental Thermodynamic Properties. *Journal of Chemical and Engineering Data*, 50, 546–550, 2005.

[16] Hukkerikar, A.S.; Sarup, B.; Ten Kate, A.; Abildskov, J.; Sin, G.; Gani, R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria*, 321, 25–43, 2012.

[17] Bendig, M. *Integration of Organic Rankine Cycles for Waste Heat Recovery in Industrial Processes*. Ph.D. thesis, Citeseer, 2015.

[18] Santos-Rodriguez, M.M.; Flores-Tlacuahuac, A.; Zavala, V.M. A stochastic optimization approach for the design of organic fluid mixtures for low-temperature heat recovery. *Applied Energy*, 198, 145–159, 2017.

[19] Joint Committee for Guides in Metrology. International vocabulary of metrology — Basic and general concepts and associated terms (VIM). Tech. rep., 2008.

[20] Sin, G.; Gernaey, K.V.; Eliasson Lantz, A. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. *Biotechnology progress*, 25, 1043–1053, 2009.

[21] Hukkerikar, A.S.; Kalakul, S.; Sarup, B.; Young, D.M.; Sin, G.; Gani, R. Estimation of environment-related properties of chemicals for design of sustainable processes: development of group-contribution+ (GC+) property models and uncertainty analysis. *Journal of chemical information and modeling*, 52, 11, 2823–39, 2012.

[22] Feistel, R.; Lovell-Smith, J.W.; Saunders, P.; Seitz, S. Uncertainty of empirical correlation equations. *Metrologia*, 53, 4, 1079–1090, 2016.

[23] Saltelli, A.; Annoni, P.; Azzini, I.; Campolongo, F.; Ratto, M.; Tarantola, S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181, 2, 259–270, 2010.

[24] Ng, L.Y.; Chemmangattuvalappil, N.G.; Ng, D.K. *Optimal Chemical Product Design via Fuzzy Optimisation based Inverse Design Techniques*, vol. 33. Elsevier, 2014.

[25] Papadopoulos, A.I.; Stijepovic, M.; Linke, P.; Seferlis, P.; Voutetakis, S. Toward Optimum Working Fluid Mixtures for Organic Rankine Cycles using Molecular Design and Sensitivity Analysis. *Industrial & Engineering Chemistry Research*, 52, 34, 12116–12133, 2013.

[26] Mavrou, P.; Papadopoulos, A.I.; Seferlis, P.; Linke, P.; Voutetakis, S. Selection of working fluid mixtures for flexible Organic Rankine Cycles under operating variability through a systematic nonlinear sensitivity analysis approach. *Applied Thermal Engineering*, 89, 1054–1067, 2015.

[27] Vidal, M.; Rogers, W.J.; Holste, J.C.; Mannan, M.S. A review of estimation methods for flash points and flammability limits. *Process Safety Progress*, 23, 47–55, 2004.

[28] Rowley, J.R.; Rowley, R.L.; Wilding, W.V. Estimation of the lower flammability limit of organic compounds as a function of temperature. *Journal of hazardous materials*, 186, 1, 551–557, 2011.

[29] Frutiger, J.; Marcarie, C.; Abildskov, J.; Sin, G. Group-contribution based property estimation and uncertainty analysis for flammability-related properties. *Journal of Hazardous Materials*, 318, 783–793, 2016.

[30] Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 183-184, 183–208, 2001.

[31] Rowley, R.L.; Wilding, W.V.; Oscarson, L., J.; Knotts, T.A.; Giles, N.F. DIPPR® Data Compilation of Pure Chemical Properties, Design Institute for Physical Properties, AIChE, New York, NY. 2014.

[32] Frenkel, M.; Chirico, R.D.; Diky, V.; Yan, X.; Dong, Q.; Muzny, C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. *Journal of chemical information and modeling*, 45, 816–838, 2005.

[33] Diky, V.; Chirico, R.D.; Muzny, C.D.; Kazakov, A.F.; Kroenlein, K.; Magee, J.W.; Abdulagatov, I.; Won Kang, J.; Gani, R.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 8. Properties of Material Streams and Solvent Design. *Chemical Information and Modeling*, 53, 249–266, 2013.

[34] Westhaus, U.; Dröge, T.; Sass, R. Detherm®—a thermophysical property database. *Fluid phase equilibria*, 158, 429–435, 1999.

[35] Molnarne, M.; Möller, W. Chemsafe–eine datenbank für bewertete sicherheitstechnische kenngrößen. *Chemie Ingenieur Technik*, 81, 1-2, 45–47, 2009.

[36] Pierotti, G.J.; Deal, C.H.; Derr, E.L. Activity Coefficients and Molecular Structure. *Industrial & Engineering Chemistry*, 51, 95–102, 1959.

[37] Huber, P.J. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35, 73–101, 1964.

[38] Seber, G.; Wild, C. *Nonlinear Regression*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 1989.

[39] Wassermann, L. *All of Nonparametric Statistics*. Springer, Berlin, 2006.

[40] Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11, 431–441, 1963.

[41] Coleman, T.F. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds ". *Mathematical Programming*, 67, 189–224, 1994.

[42] Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9, 112–147, 1998.

[43] Audet, C.; Dennis, J.E. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13, 889–903, 2003.

[44] Sin, G.; Gernaey, K.V.; Neumann, M.B.; van Loosdrecht, M.C.M.; Gujer, W. Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty. *Water Research*, 45, 639–651, 2011.

[45] Efron, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1–26, 1979.

[46] Brun, R.; Kühni, M.; Siegrist, H.; Gujer, W.; Reichert, P. Practical identifiability of ASM2d parameters - Systematic selection and tuning of parameter subsets. *Water Research*, 36, 4113–4127, 2002.

[47] Sin, G.; Vanrolleghem, P.a. Extensions to modeling aerobic carbon degradation using combined respirometric-titrimetric measurements in view of activated sludge model calibration. *Water Research*, 41, 3345–3358, 2007.

[48] Homberg, A. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. *Mathematical Bioscience*, 62, 23–43, 1982.

[49] Baltes, M.; Schneider, R.; Reuss, M. Optimal experimental design for parameter estimation in unstructured growth models. *Biotechnolgoy progress*, 10, 480–488, 1994.

[50] Kontogeorgis, G.M.; Folas, G.K. *Industrial Applications Thermodynamic Models for Industrial Applications From Classical and Advanced*. John Wiley & Son, West Sussex, UK, 2010.

[51] Lemmon, E.; Huber, M.; McLinden, M. Reference fluid thermodynamic and transport properties-REFPROP, standard reference database 23, version 8.0, National Institute of Standard and Technology;. 2007.

[52] Bell, I.; Wronski, J.; Quoilin, S.; Lemort, V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. *Industrial & Engineering Chemistry Research*, 53, 2498–2508, 2014.

[53] Peng, D.Y.; Robinson, D.B. A New Two-Constant Equation of State. *Industrial & Engineering Chemistry Fundamentals*, 15, 1, 59–64, 1976.

[54] Soave, G. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chemical Engineering Science*, 27, 6, 1197–1203, 1972.

[55] Gross, J.; Sadowski, G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Industrial & Engineering Chemistry Research*, 40, 4, 1244–1260, 2001.

[56] Mokbel, I.; Rauzy, E.; H., L.; C., B.; J., J. Vapor pressures of 12 alkylcyclohexanes, cyclopentane, butylcyclopentane and trans-decahydronaphthalene down to 0.5 Pa. Experimental results, correlation and prediction by an equation of state. *Fluid Phase Equilibria*, 108, 103–120, 1995.

[57] Garcia Baonza, V.; Caceres Alonso, M.; Nunez Delgado, J. Study of the Equation of State of Cyclopentane from 193 to 298 K and Pressures up to 104 MPa. *The Journal of Physical Chemistry*, 96, 1859–1868, 1992.

[58] Iman, R.; Conover, W. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics Part B-simulation and Computation*, 11, 3, 311–334, 1982.

[59] Helton, J.C.; Davis, F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, 81, 1, 23–69, 2003.

[60] Roskosch, D.; Atakan, B. Reverse engineering of fluid selection for thermodynamic cycles with cubic equations of state, using a compression heat pump as example. *Energy*, 81, 202–212, 2015.

[61] Sobol', I.M.; Kucherenko, S. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79, 10, 3009–3017, 2009.

[62] Halton, J.H. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2, 84–90, 1960.

[63] Frutiger, J.; Abildskov, J.; Sin, G. Global sensitivity analysis of computer-aided molecular design problem for the development of novel working fluids for power cycles. In *26th European Symposium on Computer Aided Process Engineering*, pp. 283–288. 2016.

[64] Kucherenko, S.; Rodriguez-Fernandez, M.; Pantelides, C.; Shah, N. Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering and System Safety*, 94, 1135–1148, 2009.

[65] Lampe, M.; Stavrou, M.; Bücker, H.M.; Gross, J.; Bardow, A. Simultaneous Optimization of Working Fluid and Process for Organic Rankine Cycles Using PC-SAFT. *Industrial & Engineering Chemistry Research*, 53, 21, 8821–8830, 2014.

[66] Walter, R. *Einführung in die Analysis*. Walter de Gruyter, 2007.

[67] MATLAB 2015b, The MathWorks, Inc., Natick, Massachusetts, United States.

[68] Crowl, D.A.; Louvar, J.F. Definitions of fires and explosions. In *Chemical process safety*, pp. 241–243. Prentice Hall International Series in the Physical and Chemical Engineering Sciences, Boston, 2011.

[69] Gharagheizi, F.; Mirkhani, S.A.; Tofangchi Mahyari, A.R. Prediction of standard enthalpy of combustion of pure compounds using a very accurate group-contribution-based method. *Energy Fuels*, 25, 2651–2654, 2011.

[70] Zabetakis, M.; Lambiris, S.; Scott, G. Flame temperatures of limit mixtures. In *7th Symposium (International) on Combustion*, p. 484. 1959.

[71] Rowley, J.R.; Rowley, R.L.; Wilding, W.V. Experimental determination and re-examination of the effect of initial temperature on the lower flammability limit of pure liquids. *Journal of Chemical and Engineering Data*, 55, 3063–3067, 2010.

[72] Coward, H.; Jones, G. Limits of flammability of gases and vapors. Tech. rep., Washington DC, 1952.

[73] Span, R. *Multiparameter equations of state: an accurate source of thermodynamic property data*. Springer Science & Business Media, 2010.

[74] Frutiger, J.; Bell, I.; O'Connell, J.P.; Kroenlein, K.; Abildskov, J.; Sin, G. Uncertainty assessment of equations of state with application to an organic Rankine cycle. *Molecular Physics*, 115, 2017.

[75] Aly, F.; Lee, L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. *Fluid Phase Equilibria*, 6, 3-4, 169–179, 1981.

[76] Redlich, O.; Kwong, J.N.S. On the thermodynamics of solutions; an equation of state; fugacities of gaseous solutions. *Chemical reviews*, 44, 233–244, 1949.

[77] Daubert, T.E. Vapor - Liquid Critical Properties of Elements and Compounds . 5 . Branched Alkanes and Cycloalkanes. *Journal of Chemical & Engineering Data*, 41, 365–372, 1996.

[78] Ambrose, D.; Patel, N.C. The correlation pressures and estimation of vapour IV. Extrapolation of vapour pressures and estimation of critical pressures the principle of corresponding states using two reference fluids with non-spherical molecules. *Journal of Chemical Thermodynamics*, 16, 5, 459–468, 1984.

[79] Liu, W.; Meinel, D.; Wieland, C.; Spliethoff, H. Investigation of hydrofluoroolefins as potential working fluids in organic Rankine cycle for geothermal power generation. *Energy*, 67, 106–116, 2014.

[80] Fakouri Baygi, S.; Pahlavanzadeh, H. Application of the perturbed chain-SAFT equation of state for modeling CO2 solubility in aqueous monoethanolamine solutions. *Chemical Engineering Research and Design*, 93, March, 789–799, 2015.

[81] Larsen, U.; Pierobon, L.; Haglind, F.; Gabrielii, C. Design and optimisation of organic Rankine cycles for waste heat recovery in marine applications using the principles of natural selection. *Energy*, 55, 1, 803–812, 2013.

[82] Andreasen, J.G.; Larsen, U.; Knudsen, T.; Pierobon, L.; Haglind, F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic rankine cycles. *Energy*, 73, 204–213, 2014.

[83] Kennedy, J. Particle swarm optimization. *Encyclopedia of Machine Learning*, 46, 685–691, 2010.

[84] Peng, D.Y.; Robinson, D.B. A New Two-Constant Equation of State. *Industrial and Engineering Chemistry Fundamentals*, 15, 1, 59–64, 1976.

[85] Knotts, T.; Wilding, W.V.; Rowley, R.; Giles, N.; Congore, A. The DIPPR 801 Gold Standard Systems Approach to Critical Evaluation of Thermophysical Property Data. In *19th Symposium on Thermophysical Properties*. Boulder, 2015.

[86] Zühlsdorf, B.; Kjær, J.; Cignitti, S.; Madsen, C. Improving efficiency of heat pumps by use of zeotropic mixtures for different temperature glides. 2017.

[87] Kelley, C.T. *Solving nonlinear equations with Newton's method*. Society for Industrial and Applied Mathematics, Philadelphia, 2 ed., 2003.

[88] Palma-Flores, O.; Flores-Tlacuahuac, A.; Canseco-Melchor, G. Optimal molecular design of working fluids for sustainable low-temperature energy recovery. *Computers & Chemical Engineering*, 72, 334–349, 2015.

[89] Odele, O.; Macchietto, S. Computer aided molecular design: a novel method for optimal solvent selection. *Fluid Phase Equilibria*, 82, 47–54, 1993.

[90] Kirchner, J.W. *Uncertainty Analysis and Error Propagation*. University of California, Berkeley, 2001.

[91] Oehley, E. Ableitung empirischer Gleichungen fur die untere Explosionsgrenze und den Flammpunkt. *Chemie Ingenieu Technik*, 25, 399–403, 1953.

[92] Solovev, N.; Baratov, A. Lower limit of flammability of hydrocarbon–air mixtures as a function of the molecular structure of the combustible component. *Russian Journal of Physical Chemistry*, 34, 1661–1670, 1960.

[93] Shimy, A.A. Calculating flammability characteristics of hydrocarbons and alcohols. *Fire Technology*, 6, 135–139, 1970.

[94] Shebeko, Y.N.; Ivanov, A.V.; Dmitrieva, T.M. Methods of Calculating the Lower Concentration Limits of Ignition of Gases and Vapors in Air. *The Soviet Chemical Industry*, 15, 3, 311–317, 1983.

[95] Seaton, W.H. Group contribution method for predicting the lower and the upper flammable limits of vapors in air. *Journal of Hazardous Materials*, 27, 2, 169–185, 1991.

[96] Kondo, S.; Urano, Y.; Tokuhashi, K.; Takahashi, A.; Tanaka, K. Prediction of flammability of gases by using F-number analysis. *Journal of Hazardous Materials*, 82, 113–128, 2001.

[97] Albahri, T.a. Flammability characteristics of pure hydrocarbons. *Chemical Engineering Science*, 58, 16, 3629–3641, 2003.

[98] Gharagheizi, F. Quantitative Structure - Property Relationship for Prediction of the Lower Flammability Limit of Pure Compounds. *Energy & Fuels*, 22, 3037–3039, 2008.

[99] Pan, Y.; Jiang, J.; Wang, R.; Cao, H.; Cui, Y. A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine. *Journal of Hazardous Materials*, 168, 962–969, 2009.

[100] Gharagheizi, F. A new group contribution-based model for estimation of lower flammability limit of pure compounds. *Journal of hazardous materials*, 170, 2-3, 595–604, 2009.

[101] Lazzús, J.A. Neural network/particle swarm method to predict flammability limits in air of organic compounds. *Thermochimica Acta*, 512, 1-2, 150–156, 2011.

[102] Bagheri, M.; Rajabi, M.; Mirbagheri, M.; Amin, M. BPSO-MLR and ANFIS based modeling of lower flammability limit. *Journal of Loss Prevention in the Process Industries*, 25, 2, 373–382, 2012.

[103] Mendiburu, A.Z.; de Carvalho, J.A.; Coronado, C.R. Estimation of lower flammability limits of C-H compounds in air at atmospheric pressure, evaluation of temperature dependence and diluent effect. *Journal of Hazardous Materials*, 285, 409–418, 2015.

[104] Kroenlein, K.C.; Chirico, R.; Diky, V.; Bazyleva, A.; Magee, J. Thermophysical Property Reliability Issues in the Context of Automated Consumption. In *19th Symposium on Thermophysical Properties*. Boulder, CO, 2015.

[105] Jones, M.; Frutiger, J.; Abildskov, J.; Sin, G. Safeprops: A Software for Fast and Reliable Estimation of Safety and Environmental Properties for Organic Compounds. In *AIChE Annual Meeting, , San Francisco. USA.* 2016.

[106] Leekumjorn, S.; Krejbjerg, K. Phase behavior of reservoir fluids: Comparisons of PC-SAFT and cubic EOS simulations. *Fluid Phase Equilibria*, 359, 17–23, 2013.

[107] Forster, P.; Ramaswamy, V.; Artaxo, P.; Berntsen, T.; Betts, R.; Fahey, D.W.; Haywood, J.; Lean, J.; Lowe, D.C.; Myhre, G.; *et al.* Changes in atmospheric constituents and in radiative forcing. chapter 2. In *Climate Change 2007. The Physical Science Basis*. 2007.

[108] Designation, N. Safety classification of refrigerants. *ANSI/ASHRAE Standard*, pp. 34–1992, 2010.

[109] March, J. *Advanced organic chemistry: reactions, mechanisms, and structure*. John Wiley & Sons,, 1992.

[110] Gani, R.; Zhang, L.; Kalakul, S.; Cignitti, S. *Computer-Aided Molecular Design and Property Prediction*, vol. 39. Elsevier, 2017.

Part II

Journal articles and peer reviewed conference proceedings

Paper A

A comprehensive methodology for development, parameter Estimation, and uncertainty analysis of group contribution based property models - An application to the heat of combustion

Jérôme Frutiger, Camille Marcarie, Jens Abildskov, Gürkan Sin

Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

published in Journal of Chemical and Engineering Data (ISSN: 0021-9568), 61, 1, pages: 602-613, 2016.
DOI: http://dx.doi.org/10.1021/acs.jced.5b00750

Abstract

A rigorous methodology is developed that addresses numerical and statistical issues when developing group contribution (GC) based property models such as regression methods, optimization algorithms, performance statistics, outlier treatment, parameter identifiability and uncertainty of the prediction. The methodology is evaluated through development of a GC method for prediction of the heat of combustion (ΔH_c°) for pure components. The results showed that robust regression lead to best performance statistics for parameter estimation. Bootstrap method is found a valid alternative to calculate parameter estimation errors when underlying distribution of residuals is unknown. Many parameters (first, second, third order groups contributions) are found unidentifiable from the typically available data, with large estimation error bounds and significant correlation. Due to this poor parameter identifiability issues, reporting of the 95%-confidence intervals of the predicted property values should be mandatory as opposed to reporting only single value prediction, currently the norm in literature. Moreover, inclusion of higher order groups (additional parameters) does not always lead to improved prediction accuracy for the GC-models, in some cases it may even increase the prediction error (hence worse prediction accuracy). However, additional parameters do not affect calculated 95%-confidence interval. Last but not least, the newly developed GC model of the heat of combustion (ΔH_c°) shows predictions of great accuracy and quality (the most data falling within the 95% confidence intervals) and provides additional information on the uncertainty of each prediction compared to other ΔH_c° models reported in literature.

1. Introduction

When experimental values are unavailable due to cost or time constraints, there is a strong demand for generating accurate and reliable data by predictions. In the early stage of process development, when a large number of alternative processes are evaluated and ranked, property data are often estimated, especially when new or alternative products or processes are analysed [1]. Thus, property prediction models are critically important to process systems engineering, e.g. process simulation, analysis and optimization as well as computer-aided molecular design (CAMD). Three main types of property prediction models are widely employed: group contribution (GC) [2], quantitative structure-property relationship (QSPR) [3] and *ab initio* quantum mechanics based methods [4].

GC based prediction of pure component properties uses a function of structurally dependent parameters. The best known GC methods are those of Joback and Reid [5], Lydersen [6], Klincewicz and Reid [7], Constantinou/Gani [8] and Marrero/Gani [2]. Compared to *ab initio* procedures, GC methods have a simpler model structure, a wider application range and are computationally less demanding. The advantage of the GC approach compared to quantitative structure property relationship (QSPR) or prediction based on artificial neural networks (ANN) is that the model structure does not depend on the data set [9]. This means that GC

models are likely to be more reliable for predicting properties of compounds not included in the original data set used for model building. The idea of a property function common to all species is in line with Pitzer´s corresponding states principle [10], often shown to be nearly valid for fluid properties.

In GC model development, the key task is estimation of group contributions using experimental data. Systematic reporting of uncertainty for experimental values is widely used [11]. Hence, assessing uncertainty of both estimated parameters and predicted properties is appropriate, but this issue has nevertheless traditionally not been systematically reported. While the importance of uncertainty analysis has been recognized in the literature (Whiting [12], Larsen [13], Klotz and Mathias [14], Hajipourt and Satyro [15], Maranas [16], Yan [17], Verevkin [18]), the quantification of the source of uncertainties itself (e.g. property prediction errors associated with any property models) has not received much attention. For example, Whiting [12] investigated the effects of uncertainties in thermodynamic data and models on process calculations, Larsen [13] suggested methods to analyse the data quality for chemical process design and Klotz and Mathias [14] compared van der Waals (vdW) equations of state (EOS) for specific properties. Furthermore, Hajipour and Satyro [15] illustrated the effect of uncertainty of models for critical constants and acentric factor and Maranas [16] performed an uncertainty analysis on optimization calculations involved in computer aided molecular design studies. Yan [17] compared the reliability of a variety of group contribution methods in predicting critical temperatures of organic compounds by analysing the respective average absolute deviation. Verevkin et al. [18] proposed a new group-contribution approach involving systematic corrections for 1,4-nonbonded carbon-carbon and carbon-oxygen interactions. The authors considered uncertainties of predicted values. However, their modification of the covariance matrix calculation seems non-standard, as it is not based on known statistical theories for parameter estimation [19], and its generalization may not be straightforward.

Recently, the Marrero/Gani group contribution method (MG method) was used by Hukkerikar et al. [20] to estimate thermo-physical properties (e.g. flash point) of pure components. Hukkerikar et al. performed a GC parameter estimation based on maximum likelihood theory, an uncertainty analysis based on the parameter covariance matrix and performance criteria to assess the quality. In addition to Hukkerikar et al. there is a need for a comprehensive methodology that includes

- Formulation of parameter estimation problem (e.g. weighted least squares, ordinary least squares, robust regression)
- Performance of optimization algorithms used to locate minima of the objective function used for parameter estimation
- Additional alternative uncertainty analysis method
- Assessment of parameter estimation errors and of property model prediction errors

- Method to identify outliers and data pre-treatment
- Analysis of the source of uncertainty
- Effects of additional GC-factors on prediction and uncertainty

We aim at a methodology to perform a comprehensive and step-by-step assessment and solution of the above mentioned challenges involved in developing GC-based property models. We demonstrate the methodology by developing a new GC model for the heat of combustion (ΔH_c°) based on the MG groups, employing molecular structural information at different levels.

The heat of combustion ΔH_c° provides important information in risk assessment in order to quantify the stabilities of chemical compounds. Furthermore the values are required when considering the thermal efficiency of process equipment in particular where either heat or power is produced. ΔH_c° is defined as the enthalpy increase of a chemical compound while undergoing an oxidation to defined combustion products at a temperature of 298.15 K and pressure of 1 atm [21].

There are a number of GC-based methods for the estimation of ΔH_c° in the literature. Cardozo [22] estimated enthalpies of combustion by developing correction factors from an equivalent alkane chain length and then utilized these factors along with simple relations developed for n-alkanes. Seaton and Harrison [23] proposed a method based on the original Benson's methods that had been used for the prediction of enthalpy of formation. Both Cardozo as well as Seaton and Harrison did not provide information on accuracy and uncertainty of their respective models. Hshieh et al. [24] developed an empirical model to estimate the heat of combustion. However, the application range is limited due to a small number of compounds taken into account for the parameter estimation.

Gharagheizi [25] developed a simple three-parameter quantitative structure-property relationship (QSPR). Cao et al. [26] suggested a model to estimate the heat of combustion based on an artificial neural network (ANN). Furthermore, Pan et al. [27] developed a four-parameter QSPR method. Recently Gharagheizi et al. [21] developed new GC model for the heat of combustion based on ANN. The latter four mentioned models showed all a high squared correlation coefficient between the experimental and the predicted data (>0.99). However, none of the studies includes a thorough uncertainty analysis of model predictions including for example the 95%-confidence interval of the prediction or the covariance matrix of the parameters. As a case study to highlight the application of rigorous methodology developed in this study, we develop a novel GC-based model for estimation of ΔH_c° as well as provide comprehensive assessment of uncertainties and model prediction accuracy including 95% confidence interval demonstrating the added value of using the systematic methodology for the development of GC –based property models.

The paper is organized as follows: (i) the overall methodology is outlined; (ii) the property model for ΔH_c° is developed; (iii) results of parameter estimation, using different regression methods, combined with outlier detection and uncertainty analysis, are presented, and; (iv) the new ΔH_c° model performance is compared with that of existing models.

2. Method

An overview of the methodology including the workflow, the data and techniques used at each step is shown in Figure 1.

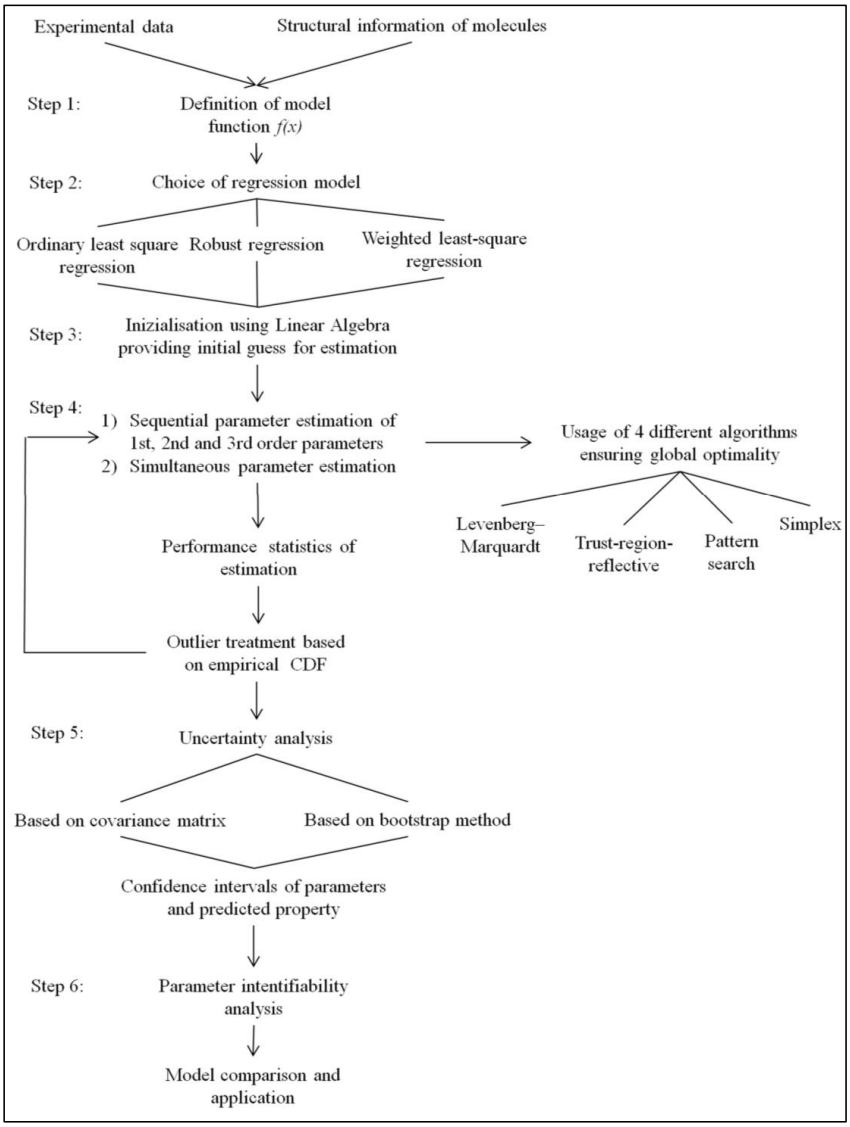


Figure 1. Overview of the methodology for development, parameter estimation and uncertainty analysis of group contribution based property models.

Detailed explanation of the tasks to perform when following the methodology is described in the following.

2.1 Property model structure definition and experimental data collection

Here the Marrero/Gani (MG) [2] method is selected for development. This method combines the contributions from a specific functional group (1st order parameters), from polyfunctional (2nd order parameters) as well as from structural groups (3rd order parameters). By using higher order parameters (2nd and 3rd), additional structural information about molecular fragments is provided. This may be useful, if the description given by 1st order groups is insufficient. The general form of the MG method is,

$$f_i(X) = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l E_l O_l \tag{1}$$

$$f(X) = T \cdot \theta \tag{2}$$

In Eq. (1) C_j is the contribution of the 1st order group of type *j* that occurs N_j times whereas D_k is the contribution of the 2nd order group of type *k* that occurs M_k times in the molecular structure of a pure component. E_l is the contribution of the 3rd order group of type *l* that has O_l occurrences. The function *f(X)* is specific for a certain property *X*. The parameters can be collected in the vector θ and the occurrences of the groups can be depicted in the matrix *T* as shown in Eq. (2). As an example, the different GC-factors of 1,2-Dichloro-4-nitrobenzene and Adiponitrile are visualized in Figure 2.

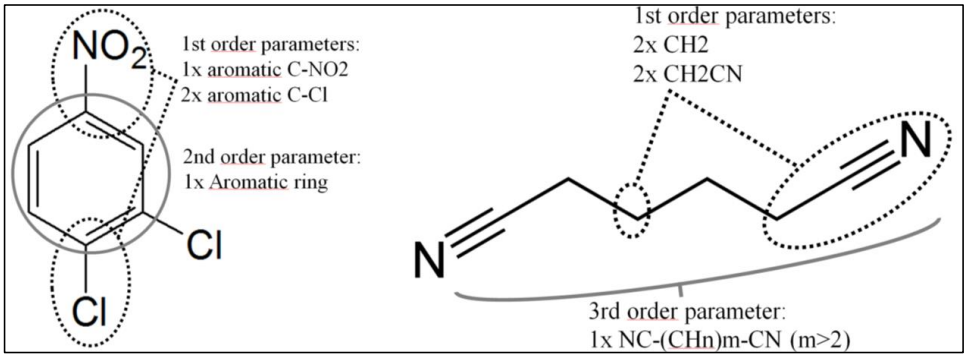


Figure 2. Example of Marrero/Gani group contribution factors of 1,2-Dichloro-4-nitrobenzene and Adiponitrile.

Experimental ΔH_c° data of 794 compounds are obtained from AIChE DIPPR 801 Database [28]. A high number of experimental data points is a prerequisite in order to obtain an accurate model with a wide application range. The heat of combustion of each compound is provided in kJ/mol.

After assigning the different 1st, 2nd and 3rd order groups to the respective molecules, it is necessary to determine a model function. We seek a function of the property which is linear in the group contributions. Hence, a suggestion for the property function is obtained by generating plots of various classes of pure components versus their increasing carbon number in homologue series as already shown by Pierotti et al. [29]. A selection of classes of compounds is shown in figure 3. From these plots, a linear function is deemed as appropriate model function for the ΔH_c° property and shown in Eq. (3), where $\Delta H_{c\,const}^\circ$ is a universal constant.

$$f(X) = f(\Delta H_c^\circ) = \Delta H_c^\circ - \Delta H_{c\,const}^\circ \tag{3}$$

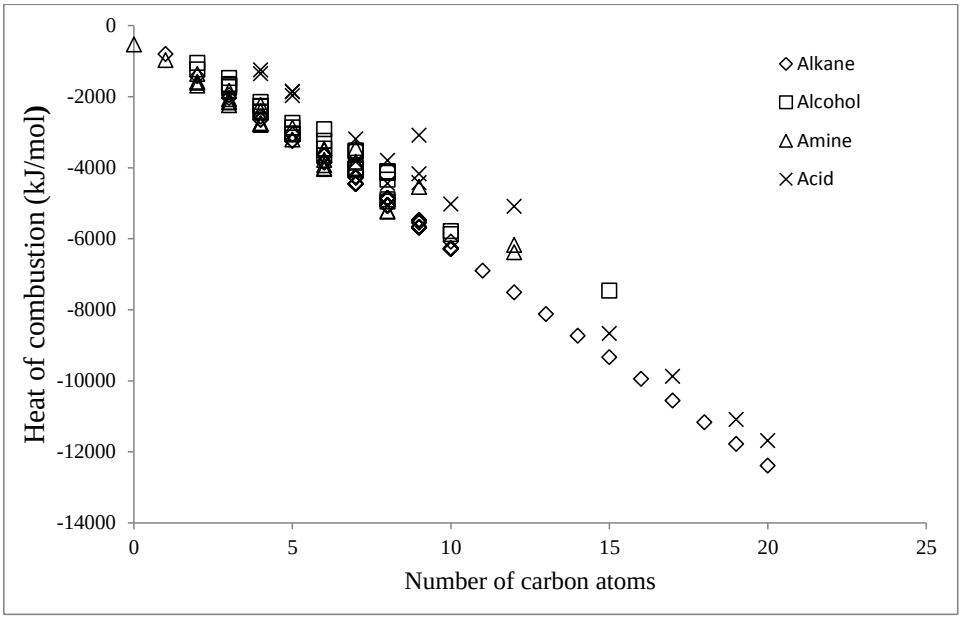


Figure 3. Graphical analysis of number of carbon atoms versus property to infer about a proper model function: (y-axis) heat of combustion ΔH_c° of a selection of pure components, (x-axis) carbon number of pure components in increasing order.

2.2. Choice of regression method

Three regression models are investigated for the use in parameter estimation in group contribution model development.

- Ordinary nonlinear least squares regression

- Robust regression

- Weighted nonlinear least squares regression

Ordinary nonlinear least squares regression is the most commonly used method for parameter estimation. The ordinary least squares regression minimizes the squares of the difference between the experimental property value y^{exp} and the predicted property value y^{pred}, i.e. the residuals, in order to get the parameter estimates θ^*,

$$\theta^* = arg\,min \quad \sum_i (y_i^{exp} - y_i^{pred})^2 \tag{4}$$

For the case of ΔH_c° y^{pred} is defined by combining Eq. (1) and (3), see Eq. (5). Each data point has equal weight (unity) [19],

$$y^{pred} = \Delta H^\circ_{c\,const} + \sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l \tag{5}$$

Ordinary least squares regression assumes that the errors are ideally independently distributed and uncorrelated, following a Gaussian distribution with a mean value of zero and a constant variance [19]. While these assumptions are made, in practice their validity is rarely checked. This is the motivation for using a bootstrap method as outlined below.

In robust regression each residual is weighted by a certain factor w_i [19]. Here the Cauchy weight is used, placing high weights on small residuals and small weights on large residuals (see Eq. (6) and (7)). The weights are updated recursively. In this way the influence of data points producing large residuals (not following the model), i.e. potential outliers, is decreased. Another intrinsic property of robust regression is that a common variance of all data points is not assumed [30].

$$\theta^* = arg\,min \quad \sum_i w_i \cdot (y_i^{exp} - y_i^{pred})^2 \tag{6}$$

$$w_i = \frac{1}{1 + (y_i^{exp} - y_i^{pred})^2} \tag{7}$$

Weighted non-linear least squares regression uses the variance V_i of the measurement error to weight the data as shown in Eq. (8) [31]. Data points with a high variance are considered to be less reliable and hence their influence on the objective function is reduced. The variance of errors of the present experimental ΔH_c° measurements are obtained from the AIChE DIPPR 801 Database [28].

$$\theta^* = arg\,min \quad \sum_i V_i^{-1} \cdot (y^{exp} - y^{pred})^2 \tag{8}$$

$$V_i = \sigma_i^2 \tag{9}$$

In Eq. (9) σ_i is the standard deviation of the respective measurement error.

2.3. Initialization using linear algebra and sequential parameter estimation

The universal constant as well as the GC factors are (a priori) unknown. A first guess $\hat{\theta}$ for the parameter estimate is provided using linear algebra according to Eq. (10),

$$\hat{\theta} = (T^{tr} \cdot T)^{-1} \cdot T^{tr} \cdot f(X) \tag{10}$$

A value for the constant $\Delta H^{\circ}_{c\,const}$ is assumed in order to calculate the first guess for the parameters from ΔH°_{c} data and the occurrence matrix T. This offers a unique solution existing without iterations.

2.4. Sequential and simultaneous parameter estimation and verification of global optimality

Afterwards the universal constants as well as the 1st, 2nd and 3rd order parameters are estimated separately and sequentially applying the non-linear regression model from the previous step. The solution of Eq. (10) is used as input for the sequential parameter estimation in the next step.

The result of the sequential estimation serves as initial guess for the simultaneous parameter estimation algorithm, where all parameters are estimated together for the chosen regression problem. The purpose of this step is twofold: (a) integrated solution of the parameter estimation problem and (b) practical verification of global optimality of the parameter estimation solution. In order to test that the global minimum of the least-squares regression has been achieved, a practical approach is followed, in which 4 different optimization algorithms are applied.

Derivative based: - Levenberg–Marquardt algorithm [32]

- Trust-region reflective algorithm [33]

Non-derivative based: - Simplex algorithm [34]

- Pattern search optimization [35]

The Levenberg-Marquardt as well as the Trust-region reflective algorithm are based on the method of steepest descent and the line search approach. They differ in the solution of the quadratic subproblems [36]. Both algorithms are commonly known as computationally very fast compared to non-derivative based algorithms. However, if the parameter number is high and the parameters are a priori unknown (as in developing GC models), it is suggested to additionally use robust non-derivative based algorithm such as simplex and pattern search [34].

Statistical performance indicators for parameter estimation. Performance of the parameter estimates is quantified by a variety of statistics in order to obtain a broad set of measures. Hukkerikar et al. [20] adopted the following statistics:

- Sum of squared errors between the experimental and predicted data,

$$SSE = \sum_i (y_i^{exp} - y_i^{pred})^2 \tag{11}$$

- Standard deviation, *SD*, measures the spread of the data about the mean value μ_y,

$$SD = \frac{1}{N} \cdot \sqrt{\sum_j (y_j^{pred} - y_j^{exp})^2} \tag{12}$$

- R^2 between the experimental and the predicted values suggests the quality of the model fit by assessing linear correlation,

$$R^2 = \frac{\sum_j (y_j^{exp} - y_j^{pred})^2}{\sum_j (y_j^{exp} - \mu_y)^2} \tag{13}$$

R^2 close to 1 indicates that the experimental data used in the regression have been fitted to a good accuracy.

- Average absolute deviation (AAD) is the measure of deviation of predicted property values from the experimentally measured property values,

$$AAD = \frac{1}{N} \sum_j |y_j^{exp} - y_j^{pred}| \tag{14}$$

- Average relative error *ARE* provides an average of relative error calculated with respect to the experimentally measured property values,

$$ARE = \frac{1}{N} \sum_j \frac{(y_j^{exp} - y_j^{pred})}{y_j^{exp}} \tag{15}$$

- The percentage of the experimental data-points P_{rc} represents the fraction of data found within ± 25% relative error range respectively.

In addition to the above suggested performance statistics, the rank correlation coefficient ρ^2, is proposed,

$$\rho^2 = (1 - \frac{6 \sum_i^n (y_i^{pred} - y_i^{exp})^2}{n(n^2-1)})^2 \tag{16}$$

Similar to the R^2, the rank correlation coefficient ρ^2 measures the quality of the model fit. A value near unity is desired. An advantage of ρ^2 is that it is more suitable to assess monotonically increasing nonlinear functions which is the nature of ranked property [37].

The classical parameter estimation problem assumes that the error of the data is normally distributed. In addition to the above statistical performance indicators suggested by Hukkerikar et al. [20], different probability plots of the residual errors are considered to test if the underlying assumptions are valid:

1. Normal probability plot: Illustrates sequential departure from Gaussian normality, hence how closely the errors follow normal distribution.
2. Cauchy probability plot: Illustrates how well the errors follow a potential Cauchy distribution, which is better suited to describe residual distributions deviating from normal distribution due to in particular long tails (residuals distribution obtained from property prediction models mostly have long tails as observed in Hukkerikar et al. [20]). The Cauchy distribution is defined as in Eq. (17) [38],

$$F_{cauchy}(x) = \frac{1}{\pi(1+x^2)}. \tag{17}$$

2.5. Outlier treatment based on empirical cumulative distribution

The GC parameter estimation can be strongly influenced by outliers from the model structure. Although principles for their detection and deletion are well known, in property modeling literature it is uncommon to see an explicit account of a systematic treatment of outliers. In engineering applications usually a normal distribution of data is assumed to be followed and residuals beyond 2-3 standard deviations are considered to be outliers. Here outliers are detected based on the empirical cumulative distribution function (CDF) of the residuals between experimental and predicted values. This methodology was suggested by Frutiger et al. [39] for the identification of outliers in group contribution models, exemplified for the upper flammability limit UFL and compared to outlier detection based on Cook's distance and normal cumulative distribution.

The empirical CDF is a step function that increases by 1/n in every data point, where n is the number of data points. In this way, it seeks to estimate the true underlying distribution function of residuals and thereby improve the detection of outliers. It does not assume that residuals follow a normal distribution (or any other distribution function a priori), as e.g. the approach suggested by Ferguson [40]. This can be an advantage if the probability plots show great deviations from Gaussian normality. Data points that lie below the 2.5% or above the 97.5% probability levels which corresponds to 2-sigma deviation in normal distribution, are taken to be outliers.

Figure 4 shows the empirical CDF of the parameter estimation using ordinary non-linear least squares regression and Levenberg-Marquardt algorithm.

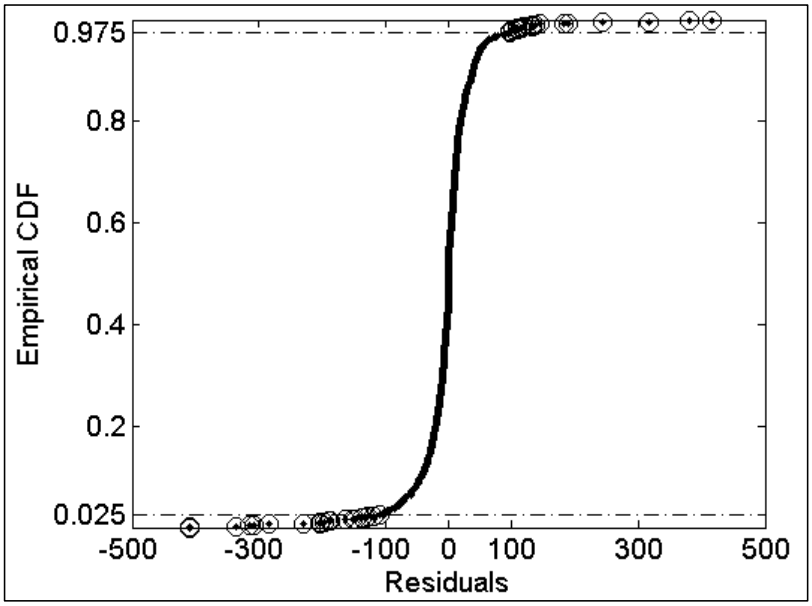


Figure 4. Empirical CDF of the residuals obtained from the parameter estimation using ordinary non-linear least squares regression and Levenberg-Marquardt algorithm. Below a probability of 0.025 and above 0.975 the data points are considered to be outliers.

3. Uncertainty of parameter estimation and property prediction

3.1. Uncertainty analysis based on linear error propagation using parameter covariance matrix

The underlying assumption of this method for uncertainty analysis method is that the measurement errors are ideally and independently distributed and defined by a Gaussian distribution white noise (normal distribution with zero mean and unit standard deviation).

The uncertainty of the parameter estimates is based on the asymptotic approximation of the covariance matrix, $COV(\theta^*)$ of parameter estimators [19], [41]

$$COV(\theta^*) = \frac{SSE}{n-p}(J(\theta^*)^T J(\theta^*))^{-1} \tag{18}$$

In Eq.(18) *SSE* is the minimum sum of squared errors obtained from the least-squares parameter estimation method, *n* is the number of data points and *p* the number of parameters. The Jacobian *J* is the local sensitivity of the property model *f* with respect to the parameter values θ^*. The corresponding elements of the parameter correlation matrix can be obtained by

$$Corr(\theta_i^*, \theta_j^*) = \frac{COV(\theta_i^*, \theta_j^*)}{Var(\theta_i^*)Var(\theta_j^*)} \tag{19}$$

In Eq. (19) $COV(\theta_i^*, \theta_j^*)$ is the respective element of θ_i^* and θ_j^* of the covariance matrix and $Var(\theta_i^*)$ and $Var(\theta_j^*)$ are the variances of the respective parameters. The error on property predictions are estimated using linear error propagation in which the covariance matrix of the predictions $COV(y^{pred})$ is approximated using the Jacobian and the covariance of the parameter estimates as shown in Eq. (6),

$$COV(y^{pred}) = J(\theta^*)COV(\theta^*)J(\theta^*)^T \tag{20}$$

If the assumptions behind the model are satisfied (as ensured in previous steps) the parameter estimates will follow a student *t*-distribution, so

$$\theta_{1-\alpha}^* = \theta \pm \sqrt{diag(COV(\theta^*))} \cdot t(n-p, \alpha_t/2) \tag{21}$$

Similarly, the confidence intervals of the property predictions are given by:

$$y_{1-\alpha}^{pred} = y^{pred} \pm \sqrt{diag(J(\theta^*)COV(\theta^*)J(\theta^*)^T)} \cdot t(n-p, \alpha_t/2) \tag{22}$$

In Eq. (21) and (22) $t(n-p, \alpha_t/2)$ is the *t*-distribution value corresponding to the $\alpha_t/2$ percentile of Students t-distribution, $diag(COV(\theta^*))$ represents the diagonal elements of $COV(\theta^*)$ and $diag(J(\theta^*)COV(\theta^*)J(\theta^*)^T))$ the corresponding diagonal elements of $J(\theta^*)COV(\theta^*)J(\theta^*)^T$.

3.2. Uncertainty analysis based on bootstrap method

Using the parameter covariance matrix as described, assumes that the residuals are independent and follow normal distribution with zero mean [19]. However in practice this is rarely such (see e.g. the residual plots in Hukkerikar et al. [20]). The bootstrap method is an attempt to calculate the distributions of the errors from the data, and to use these to calculate the errors on the parameter estimation [42]. In a certain sense, the bootstrap method aims to relax the restriction to independent and identically distributed measurement errors, which is a central assumption in nonlinear least squares theory. In order to perform bootstrap method [42], first a reference parameter estimation is made, giving

$$\theta^* = arg\,min \quad \sum_i (y_i^{exp} - y_i^{pred}(\theta^*)^2) \tag{23}$$

The bootstrap method defines $\hat{F}$ as the sample probability distribution of the errors $\hat{\epsilon}$:

$$\hat{F} = mass\frac{1}{n} \quad at \quad \hat{\epsilon}_i = (y_i^{exp} - y_i^{pred}(\theta^*)) \tag{24}$$

From this the new set of errors can be obtained. The residuals are assumed to be uniformly distributed $\hat{F}$ i.e. each residual has equal probability of realization. In the next steps, new synthetic data sets are produced. The bootstrap method generates any number of synthetic data sets *($y^*(1)$; $y^*(2)$,…, $y^*(k)$)* also with *n* data points (*n* being here the total number of observations, and *k* being the total number of bootstrap samples) by using random sampling with replacement from the residuals $\hat{\epsilon}$. The procedure is simply to add the *n* bootstrap samples of residuals to the model predictions obtained using the estimated parameters in the reference step above as follows:

$$y_i^* = y_i^{pred}(\theta) + \hat{\epsilon}_i \quad where \quad \hat{\epsilon}_i \in \hat{F} \tag{25}$$

Parameter estimation is repeated using each synthesis data set *y*(k)*, which results in a new set of estimated parameters $\theta^*(k)$ and a new predicted value of $y_i^{pred*}(k)$ solving the minimization problem as formulated above. The resulting sample of estimated parameter values are plotted to graphically visualize the uncertainty in the estimated parameter values. In addition, inference statistics can be used to estimate the mean μ_{θ^*} and standard deviation $\sigma_{\theta^*}$of the distribution of the estimated parameter values. The mean value and the standard deviation of all the estimated parameter sets can be used to calculate the confidence intervals:

$$\mu_{\theta^*} = \frac{1}{n}\sum_{k=1}^{n}\theta^*(k) \tag{26}$$

$$\sigma_{\theta^*} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(\theta^*(k) - \mu_{\theta^*})^2} \tag{27}$$

In Eq. (26) and (27) *n* is the number of data points and $\theta^*(k)$ is the estimated parameter using the *k-th* synthetic data set.

3.4. Parameter identifiability

Parameter identifiability is a common issue in nonlinear regression [19] with important implications for model validation and application. Parameter identifiability is basically the issue, can the model parameters be estimated *uniquely* from a certain data set? We use the following diagnostic measures to analyze parameter identifiability in GC models:

a) The parameter estimates *must not be linearly dependent*, so the linear correlation coefficients between parameter estimates should be sufficiently low, e.g. less than 0.7 [43], [44], and
b) Parameter estimation errors (i.e. 95% confidence intervals) should be sufficiently low [45]. One obvious indication of poor parameter identifiability is a large confidence interval, e.g. relative parameter estimation error being larger than 50% [46][45].

4. Results and Discussion

4.1. Regression models and practical global optimality of parameter estimation

The performance of the applied regression models for the ΔH_c° GC method is shown in table 1. The results are depicted before and after outlier deletion, where N_{out} is the number of outliers removed. R^2, ρ^2, *SD*, *ARE*, *SSE* and *AAD* are defined above. P_{rc} represents the percentage of the experimental data points found within ± 25% relative error range respectively [20]. Figure 5 shows the prediction of ΔH_c° versus the experimental value for Robust regression and Weighted least squares regression *after* outlier deletion.

Table 1. Regression model performance statistics, the best value of the respective column is highlighted.

	R^2 *Pearson*	ρ^2 *Spearman*	N_{out}	*SD*	*AAD*	*ARE [%]*	*SSE*	P_{rc} *25%*
Ordinary least-squares *before* outlier deletion	0.99	0.99	0	76.63	30.35	1.10	$4.66 \cdot 10^6$	99.75
Robust regression *before* outlier deletion	0.99	0.99	0	87.33	21.80	0.75	$3.29 \cdot 10^5$	99.62
Weighted least squares *before* outlier deletion	0.99	0.99	0	134.39	61.25	1.82	$6.89 \cdot 10^4$	99.62
Ordinary least-squares *after* outlier deletion	0.99	0.99	40	**22.14**	14.18	0.52	$3.70 \cdot 10^5$	**100**
Robust regression *after* outlier deletion	0.99	0.99	40	23.30	**13.09**	**0.50**	$1.42 \cdot 10^5$	99.87
Weighted least squares *after* outlier deletion	0.99	0.99	40	29.64	18.37	0.57	$\mathbf{2.84 \cdot 10^3}$	**100**

Outlier deletion improves the regression performance. After outlier deletion, the results are relatively close for the three models. The best fit according to *ARE* and *AAD* was achieved by robust regression after outlier deletion. However, for robust regression *SD* is slightly higher than ordinary least squares and *SSE* is slightly higher than weighted-least squares and P_{rc} is

slightly lower compared to both of them. The regression models performed an very good fit (see Figure 5).

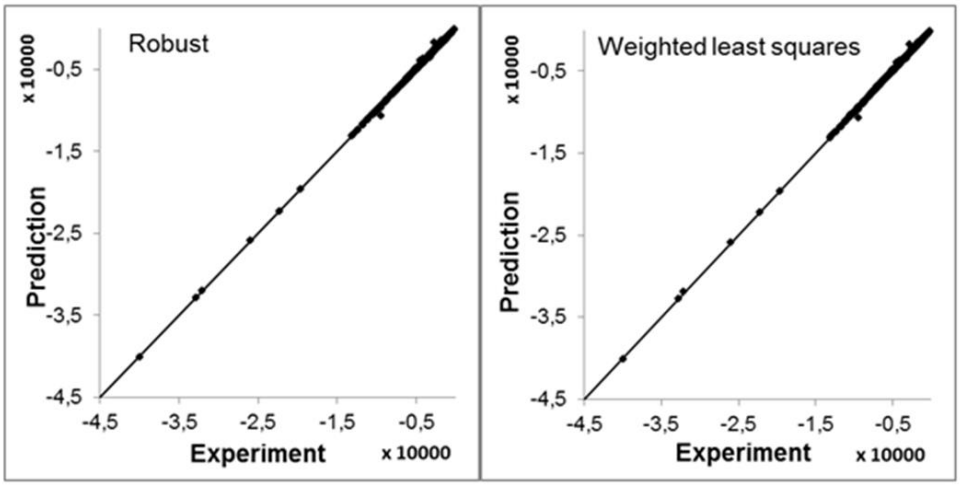


Figure 5. Prediction versus the experimental value of ΔH_c° after outlier removal for a) robust (left) and b) weighted least squares regression (right).

The reason why weighted least square performs slightly worse in terms of *SD*, *ARE* and *AAD* than robust regression can be explained as follows: The measurement error, which is the basis for the variance in the regression model, is given in percentage. Hence, large data points are often assigned a large variance and are therefore weighted less, such that the minimization of the residuals of large data points has a lower influence on the optimization. As a consequence, weighted least squares regression fits small property values much better than the large property values, whereas robust regression has no such bias. In that sense overall robust regression seems slightly favorable model for the GC parameter estimation of ΔH_c° property data.

Four separate search algorithms were used to cross-check and validate the global minimum of the solution. Table 2 shows the sum of squares errors *SSE* after the corresponding sequential and the simultaneous parameter estimation. A higher amount of parameters increases the goodness of the fit.

When comparing the final performance of the different optimization algorithms (see Table 2, final *SSE*), it can be seen that the Simplex and Trust-region reflective-algorithm lead to the best solutions, whereas *SSE* for pattern search algorithm and Levenberg-Marquart-algorithm was terminated at a higher *SSE* value. The solution found by the Simplex and Trust-region reflective- algorithms can be considered as *practically* (considering the four different search algorithms) globally optimal solution. The Levenberg-Marquart and Trust-region reflective-algorithm are strongly depending on the initial guess, since they are local search algorithms. The initial guess might have been suitable for Trust-region reflective, but not for Levenberg-

Marquart. A possible explanation why pattern search did not find the same minimum as the others could be the nature of the search algorithm. It is known to be powerful for specific classes of functions [47].

Table 2. Sum of squares errors *SSE* of the parameter estimation for different optimization algorithms using sequential (sequ.) and simultaneous (sim.) estimation.

	SSE (sequ.) 1st order	*SSE (sequ.) 1st and 2nd order*	*SSE (sequ.) 1st, 2nd and 3rd order*	***Final SSE (sim.)***
Simplex	$9.65 \cdot 10^6$	$6.32 \cdot 10^5$	$3.33 \cdot 10^6$	$\mathbf{3.93 \cdot 10^5}$
Pattern search	$1.95 \cdot 10^7$	$1.69 \cdot 10^7$	$1.30 \cdot 10^7$	$\mathbf{1.26 \cdot 10^7}$
Levenberg-Marquart	$5.95 \cdot 10^6$	$5.39 \cdot 10^6$	$5.26 \cdot 10^6$	$\mathbf{4.66 \cdot 10^6}$
Trust-region reflective	$5.31 \cdot 10^5$	$4.82 \cdot 10^5$	$4.52 \cdot 10^5$	$\mathbf{3.70 \cdot 10^5}$

4.2. Uncertainty analysis property prediction errors

Figures 6 and 7 show the experimental and the predicted values of the heat of combustion with the respective 95%-confidence interval of the prediction for every substance both for covariance-based uncertainty analysis bootstrap sampling-based methods. As an example the prediction based on parameter estimates obtained using the robust regression is shown. The compounds are ordered from lowest to highest value and given an index number respectively. The confidence intervals are individual for each compound. The trend is a narrow band along with the experimental values.

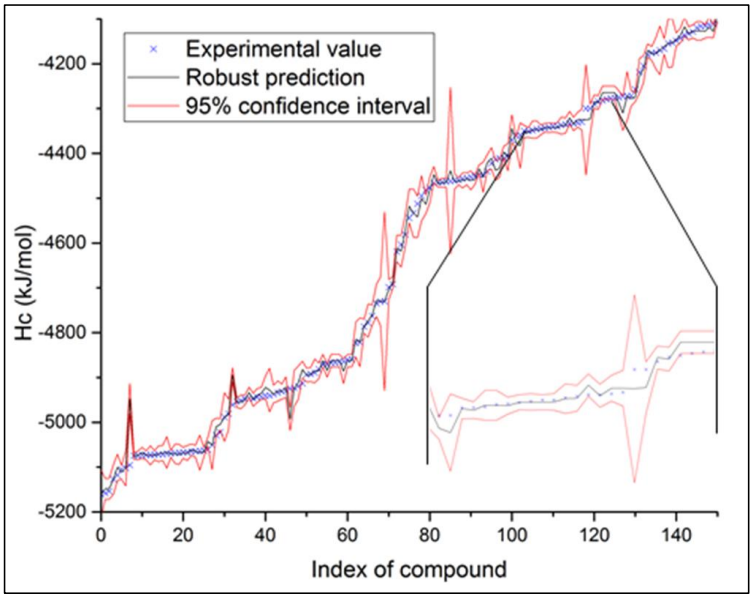


Figure 6. Experimental as well as predicted value of ΔH_c° for every compound with 95%-confidence intervals generated by covariance-based uncertainty analysis (robust regression without outliers). A section of the plot is enlarged to show the distribution of the experimental values around the prediction.

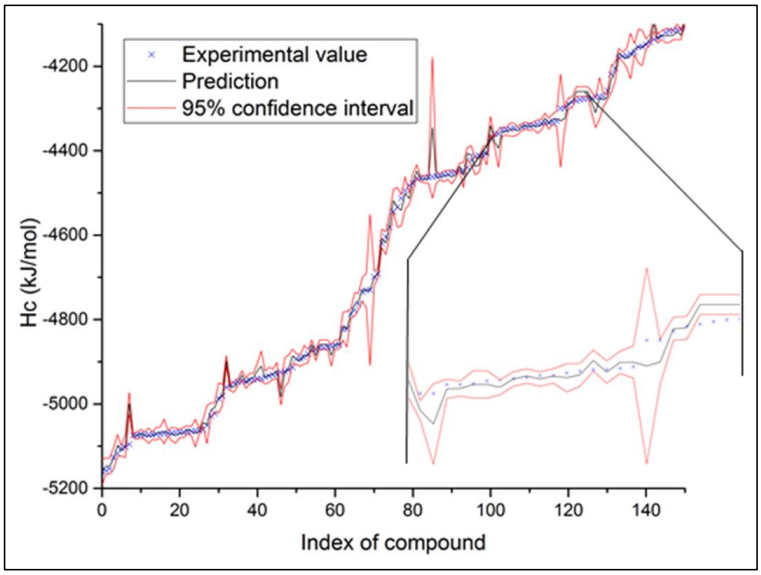


Figure 7. Experimental as well as predicted value of ΔH_c° for every compound with 95%-confidence intervals generated by bootstrap sampling-based uncertainty analysis.

Both methods (linear error propagation versus bootstrap) used for the calculation of the uncertainty of the prediction of the corresponding experimental value show a similar result, i.e. - in both methods the experimental value lies within the calculated 95% confidence intervals. Although bootstrap technique requires more model evaluations and computations compared to the linear error propagation (where only one model evaluation is needed), it has the advantage of being sampling- based, which allows non-linear error propagation.

4.3. Parameter identifiability analysis

The consideration of the 95%-confidence interval of the parameter estimates (see appendix), allows evaluating the practical identifiability of the GC factors. Although for all regression models the parameter fit was satisfying (see 3.1), there is a large number of parameters that have a large confidence interval corresponding to a relative parameter estimation error $\sigma_{\theta^*}/\mu_{\theta^*}$ being larger than 50%. For the use of ordinary least-squares regression 96 out of 235 parameters are not practically identifiable, whereas for robust regression it is 95 out of 235. 83 out of 235 parameters fail practical identifiability for weighted least-squares. However, the universal parameter $\Delta H^{\circ}_{f,const}$ is identifiable. Furthermore, almost all of the 1st order parameters (beside 3) could be identified practically compared to 2nd and 3rd order parameters where a larger part is not practically identifiable.

The practical identifiability depends on two main issues: The amount of data for the parameter estimation and the correlation between parameters.

If there is sufficient information (i.e. enough data points) to calculate the parameter estimates, the confidence interval gets smaller and hence, the parameters are practically identifiable. However, in GC parameter estimation there might be several functional groups that only occur in very few compounds. For some 3rd order parameters, there was only one compound available with a certain functional group. Hence, the 95%-confidence interval is very high and the parameters get non-identifiable

The second major source of parameter identifiability problems is high correlation (>0.7) between parameters, which can be observed in the parameter correlation matrix given in the supplementary material. The elements of the correlation matrix are directly linked to the covariance of two parameters, which is subsequently obtained from the Jacobian (see Eq. (18) and (19)). This means, if two parameters have a similar or identical sensitivity to the model output, they are highly correlated. In GC methods, correlation is intrinsically often the case, because certain functional groups can occur frequently together (depending on the data set) [48].

In many property modeling studies, practical identifiability of parameters has either not been considered or neglected. The diagnostic measures mentioned above indicate clearly that not all of the model parameters are uniquely identifiable. The first implication of this is that the estimated parameter values should not be attributed physical meaning since their values are not unique. Second, for practical application purposes, it is desirable to keep the parameters in

the model, despite their identifiability issues, because in this way the application range of the GC model is higher (the more first, second and third order group contribution parameters in the model, the more chemicals property can be predicted).

However in that case, i.e. using a model with poorly identifiable parameters, the uncertainty of the prediction (i.e. perform propagation of parameter estimation errors to the property prediction) as shown in figures 6 and 7 becomes critical. The confidence interval of property prediction provides a measure of the prediction quality (accuracy) of the model developed, which the end user can use to judge if the prediction accuracy is fit for the intended application or else a more accurate measurement needs to be done instead of using a model prediction.

4.4. Effect of addition of higher order groups on property value and uncertainty. It is valuable to analyze, what the influence of correlated parameters is on the prediction and on the uncertainty of the prediction. The results obtained in this study showed that high correlation influences the mean prediction but not the uncertainty bounds (the upper and lower 95% confidence interval). In 155 out of 794 molecules the introduction of 2nd or 3rd order groups increased the relative error between experimental and predictive values for more than 10%. This particularity is exemplified and investigated by using two compounds namely cis,trans-2,4-Hexadiene and Acrolein. The parameter correlation matrix given in Table 3, shows that the GC factors of cis,trans-2,4-Hexadiene are highly correlated in comparison to the GC factors of Acrolein. The prediction and 95%-confidence interval for the two selected substances are shown in Table 4 considers 1st order only, 1st and 2nd order as well as 1st, 2nd and 3rd order GC factors. These two examples shows that while adding more groups increases the relative error of prediction for cis,trans-2,4-Hexadiene compound (worse case), however it leads to a lower relative prediction error for Acrolein (better case). However, it does not affect the calculation of the 95%-confidence interval of the property prediction (reliable case). To understand this, we need to look back at the non-linear regression theory and parameter identifiability issues in detail.

Table 3. Parameter correlation matrices. The red color indicates a positive correlation of higher than 0.7 and the orange color indicates a negative correlation lower than -0.7.

GC-factors	$\Delta H^{\circ}_{f,const}$	"CH3	'CH=CH'	'CHn=CHm-CHp=CHk'	'CH3-CHm=CHn'
$\Delta H^{\circ}_{f,const}$	1.00				
'CH3'	-0.96	1.00			
'CH=CH'	-0.02	0.03	1.00		
'CHn=CHm-CHp=CHk'	0.02	-0.03	-0.94	1.00	
'CH3-CHm=CHn'	0.02	-0.07	-0.96	0.86	1.00

cis,trans-2,4-Hexadiene

GC-factors	$\Delta H^{\circ}_{f,const}$	'CH2=CH'	'CHO'	'CHm=CHn-CHO'
$\Delta H^{\circ}_{f,const}$	1.0			
'CH2=CH'	-0.45	1.0		
'CHO'	0.61	-0.26	1.0	
'CHm=CHn-CHO'	0.01	-0.55	0.30	1.0

Acrolein

Table 4. Prediction and 95%-confidence interval for a selection of substances comparing the usage of only 1st order GC-factors with the usage of 1st and 2nd as well as 1st, 2nd 3rd order groups.

	Relative error between prediction and experimental value			Boundary of 95%-confidence interval		
Used GC-factors	1st	1st, 2nd	1st, 2nd, 3rd	1st	1st, 2nd	1st, 2nd, 3rd
cis,trans-2,4-Hexadiene	0.024	0.038	0.029	±13.96	±13.93	±13.14
Acrolein	0.0094	0.0051	0.0051	±32.68	±32.67	±32.67

Table 5. Comparison of sample variance, s^2, as a function of increasing GC model parameters: comparison between a GC model containing only 1st order, 1st and 2nd as well as 1st, 2nd and 3rd order groups.

	Levenberg-Marquart algorithm		
Used GC-factors	1st	1st, 2nd	1st, 2nd, 3rd
SSE	531121	481983	452126
$n-p$	627	555	523
$S=\frac{SSE}{n-p}$	847	868	864

SSE is the sum of squared errors, *n* is the number of compounds for which experimental data is available and *p* the number of parameters.

In the case of Acrolein, most of the parameters are not significantly correlated and the relative error between experimental and predicted value as well as the 95%-confidence interval gets smaller by the introduction of 2nd order group. This outcome is observed for the majority (80%) of the estimated compounds considered in this work. However, cis,trans-2,4-Hexadiene shows high correlation between the parameters, in particular negative correlation. The negative correlation between the universal constant and the 1st order parameters and 1st and 2nd as well as between 1st and 3rd order groups has an influence on the prediction. The relative error increases for cis,trans-2,4-Hexadiene by the introduction of the 2nd order group. However, the uncertainty (i.e. the 95%-confidence interval) is not enlarged by the introduction of higher order group. This particularity can be understood by looking at Eq. (18) and (20) (see above). The first reason lies in the negative correlation. If two parameters are negatively correlated and have similar sensitivity to the model output (corresponding to the Jacobian $J(\theta^*)$), their uncertainties will tend to cancel [48]. The second cause is the nature of the calculation of mean sum of squared error *S=SEE/(n-p)*. Table 5 shows that this normalization factor for the covariance matrix remains constant, because the relative decrease of SSE is compensated by the corresponding increase in the number of parameters used for its estimation.

As a result, one can conclude that definition and inclusion of higher groups for a GC model may not always lead to a more accurate property prediction. At least for some chemical compounds relative prediction error will become worse due to parameter identifiability issues. This can be for GC models that have a large amount of factors to ensure a brought applicability. However, the 95%-confidence interval does not enlarge due to poor parameter identifiability. We suggest therefore that developers and users of GC models in general always

state the 95%-confidence interval, which includes information on the parameter correlation structure associated with poor parameter identifiability issues.

4.5. Comparison of different classes of compound classes

The average relative error *ARE*, the average absolute deviation and the number of compounds included for some selected classes of chemicals are shown in Table 6. The data is ordered according to the number of data points.

Table 7. Comparison of performance of different classes of chemicals.

Class	*ARE* (ΔH_c°) in %	*AAD* (ΔH_c°) in kJ/mol	No. of ΔH_c° compounds
Aromatic Compounds	0.18	10.97	104
Alkanes	0.14	7.09	103
Alkenes	0.24	8.70	65
Acids	1.04	17.29	60
Alcohols	0.41	12.70	56
Sulfur containing Compounds	0.46	11.55	44
Amines	0.70	17.91	37
Halogen containing Compounds	1.27	10.93	33
Ketones	0.52	14.72	30
Nitro-Compounds	0.51	11.13	26
Carboxylates	0.66	24.55	25
Esters	0.70	16.31	24
Ethers	0.49	13.43	20
Nitriles	0.92	13.12	18
Aldehydes	0.39	6.16	13
Pyridines	0.46	18.66	12

Overall all the major classes of chemicals, except for the Halogen containing compounds which has an ARE of 1.27%, have an ARE below 1%. In particular the model performs best for Alkanes, Aromatic Compounds and Alkenes with an *ARE* below 0.3%. This demonstrates the accuracy of the model over a great variety of chemical compounds. The classes not included in the Table 6 consist of 10 or less compounds and the corresponding results can be found in the supporting material.

4.6. Comparison of the new GC model with other property estimation models

The squared Pearson correlation coefficient R^2, average relative error *ARE*, the average absolute deviation and the number of data included of this study for the model using robust regression are compared to 5 other property prediction models in Table 6: Another group contribution (GC), quantitative structure-property relationship (QSPR), as well as artificial neural networks (ANN) for the calculation of ΔH_c°.

Table 7. Comparison of present model with existing models.

	Current study	Hshieh et al. [24] 2003	Gharageizi [25] 2008	Cao et al. [26] 2009	Pan et al. [27] 2011	Gharagheizi et al. [21], 2011
Model structure	**MG GC (robust reg.)**	Empirical Atomic Indices.	QSPR	QSPR with ANN	QSPR	ANN
R^2 Pearson	**0.99**	0.99	0.99	0.99	0.99	0.99
ARE (ΔH_c°) in %	**0.51**	3.90	3.45	-	-	0.16
AAD (ΔH_c°) in kJ/mol	**13.03**	-	-	155.32	104.13	-
No. of ΔH_c° data	**794**	75	1714	1496*	1650*	4590

*included experimental and predicted data hence it is biased.

Considering the average relative error *ARE* of ΔH_c°, the model developed in this study performs better than Hshieh et al.. Furthermore, the amount of data that is taken into account is much higher for the present model. This increases the application range of the model, since more substance from different classes of molecules have been used. In terms of *ARE* the model shows increased performance compared to Gharagheizi (2008), although the number of data points are lower. This is an indication that the parameter estimation methodology is very efficient. Cao et al. and Pan et al. have a higher absolute average error *AAE* than the new

model. Furthermore, the amount of data consists of all experimental and predicted data available in the DIPPR database which is not a proper way to perform model development and performance statistics (which should solely be based on experimental data points only). The ANN model of Gharagheizi (2011) has a lower *ARE* and more data points. ANN is a fundamentally different approach to GC models. As regards the comparison of two different approaches for heat of combustion modelling, it is important to note that in ANN approach the aim is to build the best possible model structure (i.e. how many variables, descriptors, to include). In GC-based approach, the model structure is fixed. Therefore, the aim is instead on identifying and estimating in the best possible way the parameters of the fixed model given a certain available set of measurements. Therefore, the structure of the MG GC model is much simpler compared to ANN and much easier to work with and apply in industrial applications. Furthermore the reliability of the GC model predictions have been statistically demonstrated and verified against application in practice. However, establishing the reliability and confidence of parameter estimation in ANN remains to be demonstrated. Furthermore, due to the fact that the model is predefined, new experimental values can be added to the parameter estimation without changing the model structure in GC models, while in QSPR and ANN model building need to be performed all over again.

5. Conclusion

In this study, a systematic methodology for the development, parameter estimation and uncertainty analysis of GC models was developed. The methodology was successfully applied for the development of new GC-based model with improved prediction performance statistics for the heat of combustion (ΔH_c^o). In particular, the systematically developed new model has a higher accuracy than existing GC models and is much simpler to apply than ANN models.

The following are the main conclusions from the systematic development of GC-based models:

- Concerning the regression models, robust regression showed best performance statistics.
- The bootstrap method can be considered as a valid alternative to classical uncertainty analysis (linear approximation of covariance matrix of parameter estimators) when the underlying distribution of errors is considered to be unknown or not normally distributed.
- Although GC-based models have severe parameter identifiability issues characterized by significant correlation between estimated parameters and large confidence interval, the GC-based models still can be used successfully provided that 95% confidence interval of model predictions (prediction accuracy) are also calculated and reported.
- Addition of higher order groups (additional parameters) may in certain cases increase the prediction error, but does not enlarge the uncertainty (95%-confidence interval), due to parameter correlation associated with poor parameter identifiability.
- The use of different optimization algorithms for the parameter estimation is suggested as a simple method to ensure that the practically globally optimal solution was found.

GC-based property models are highly valuable and effective tools of property predictors. To ensure accurate and reliable estimation of properties of interest, comprehensive uncertainty analysis in particular 95% confidence interval of model predictions must be performed using systematic methods as presented in this work.

References

[1] Gani R, O'Connell JP. Properties and CAPE: From present uses to future challenges. Comput Chem Eng 2001;25:3–14.

[2] Marrero J, Gani R. Group-contribution based estimation of pure component properties. Fluid Phase Equilib 2001;183–184:183–208.

[3] Bünz A., Braun B, Janowsky R. Quantitative structure–property relationships and neural networks: correlation and prediction of physical properties of pure components and mixtures

from molecular structure. Fluid Phase Equilib 1999;158–160:367–74. doi:10.1016/S0378-3812(99)00058-8.

[4] Peterson K a., Feller D, Dixon D a. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. Theor Chem Acc 2012;131:1–20. doi:10.1007/s00214-011-1079-5.

[5] Joback K, Reid R. Estimation of pure-component properties from group-contribution. Chem Eng Commun 1987;57:233–43.

[6] Lydersen AL. Estimation of critical properties of organic compounds. Coll Eng Univ Wisconsin Eng Exp Stn Rep 3, Madison, WI 1955.

[7] Klincewicz K. Estimation of critical properties with group contribution methods. AICHE J 1984;30:137–42.

[8] Constantinou L, Gani R. New group contribution method for estimating properties of pure compounds. AIChE J 1994;40:1697–709.

[9] Hukkerikar AS, Meier RJ, Sin G, Gani R. A method to estimate the enthalpy of formation of organic compounds with chemical accuracy. Fluid Phase Equilib 2013;348:23–32.

[10] Poling BE, Prausnitz JM, O'Connell JP. The Estimation of Physical Properties. Prop. Gases Liq., New York: McGraw-Hill; 2004, p. 1–9.

[11] Dong Q, Chirico RD, Yan X, Hong X, Frenkel M. Uncertainty Reporting for Experimental Thermodynamic Properties. J Chem Eng Data 2005;50:546–50.

[12] Whiting WB. Effects of Uncertainties in Thermodynamic Data and Models on Process Calculations †. J Chem Eng Data 1996;41:935–41. doi:10.1021/je9600764.

[13] Larsen AH. Data quality for process design. Fluid Phase Equilib 1986;29:47–58.

[14] Mathias P, Klotz H. Take a closer look at thermodynamic property models. Chem Eng Prog 1994;90:67–75.

[15] Hajipour S, Satyro MA. Uncertainty analysis applied to thermodynamic models and process design – 1. Pure components. Fluid Phase Equilib 2011;307:78–94. doi:10.1016/j.fluid.2011.05.014.

[16] Maranas C. Optimal molecular design under property prediction uncertainty. AICHE J 1997;43:1250–64.

[17] Yan X, Dong Q, Hong X. Reliability analysis of group-contribution methods in predicting critical temperatures of organic compounds. J Chem Eng Data 2003;48:374–80. doi:10.1021/je025596f.

[18] Verevkin SP, Emel'yanenko VN, Diky V, Muzny CD, Chirico RD, Frenkel M. New group-contribution approach to thermochemical properties of organic compounds: Hydrocarbons and oxygen-containing compounds. J Phys Chem Ref Data 2013;42:1–48. doi:10.1063/1.4815957.

[19] Seber G, Wild C. Nonlinear Regression. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1989.

[20] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.

[21] Gharagheizi F, Mirkhani SA, Tofangchi Mahyari AR. Prediction of standard enthalpy of combustion of pure compounds using a very accurate group-contribution-based method. Energy Fuels 2011;25:2651–4.

[22] Cardozo RL. Prediction of the enthalpy of combustion of organic compounds. AIChE J 1986;32:844–8. doi:10.1002/aic.690320514.

[23] Seaton WH, Harrison BK. A new general method for estimation of heats of combustion for hazard evaluation. J Loss Prev Process Ind 1990;3:311–20.

[24] Hshieh FY, Hirsch DB, Beeson HD. Predicting heats of combustion of polymers using an empirical approach. Fire Mater 2003;27:9–17. doi:10.1002/fam.815.

[25] Gharagheizi F. A simple equation for prediction of net heat of combustion of pure chemicals. Chemom Intell Lab Syst 2008;91:177–80. doi:10.1016/j.chemolab.2007.11.003.

[26] Cao HY, Jiang JC, Pan Y, Wang R, Cui Y. Prediction of the net heat of combustion of organic compounds based on atom-type electrotopological state indices. J Loss Prev Process Ind 2009;22:222–7.

[27] Pan Y, Jiang JC, Wang R, Jiang JJ. Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization. J Loss Prev Process Ind 2011;24:85–9.

[28] Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE) 2014.

[29] Pierotti GJ, Deal CH, Derr EL. Activity Coefficients and Molecular Structure. Ind Eng Chem 1959;51:95–102. doi:10.1021/ie50589a048.

[30] Huber PJ. Robust Estimation of a Location Parameter. Ann Math Stat 1964;35:73–101.

[31] Wassermann L. All of Nonparametric Statistics. Berlin: Springer; 2006.

[32] Marquardt D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J Soc Ind Appl Math 1963;11:431–441.

[33] Coleman TF. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds ". Math Program 1994;67:189–224.

[34] Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the nelder–mead simplex method in low dimensions. SIAM J Optim 1998;9:112–47.

[35] Audet C, Dennis JE. Analysis of generalized pattern searches. SIAM J Optim 2003;13:889–903.

[36] Byrd RH, Schnabel RB, Shultz GA. A Trust Region Algorithm for Nonlinearly Constrained Optimization. SIAM J Numer Anal 1987;24:1152–1170.

[37] Louangrath PI. Correlation Coefficient According to Data Classification. SSRN Electron J 2014:1–28. doi:10.2139/ssrn.2417910.

[38] Ferguson TS. Maximum Likelihood Estimates of the Parameters of the Cauchy Distribution for Samples of Size 3 and 4. J Am Stat Assoc 1978;73:211–3.

[39] Frutiger J, Abildskov J, Sin G. Outlier treatment for improving parameter estimation of group contribution based models for upper flammability limit. In: Gernaey K V., Huusom JK, Gani R, editors. 12th Int. Symp. Process Syst. Eng. 25th Eur. Symp. Comput. Aided Process Eng., Copenhagen: 2015.

[40] Ferguson TS. On the rejection of outliers. Proc Berkeley Symp Math Stat Probab 1961;1:253–87.

[41] Sin G, Gernaey K V., Neumann MB, van Loosdrecht MCM, Gujer W. Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty. Water Res 2011;45:639–51. doi:10.1016/j.watres.2010.08.025.

[42] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.

[43] Brun R, Kühni M, Siegrist H, Gujer W, Reichert P. Practical identifiability of ASM2d parameters - Systematic selection and tuning of parameter subsets. Water Res 2002;36:4113–27. doi:10.1016/S0043-1354(02)00104-5.

[44] Sin G, Vanrolleghem P a. Extensions to modeling aerobic carbon degradation using combined respirometric-titrimetric measurements in view of activated sludge model calibration. Water Res 2007;41:3345–58. doi:10.1016/j.watres.2007.03.029.

[45] Homberg A. On the practical identifiability of microbial growth models incorporating

Michaelis-Menten type nonlinearities. Math Biosci 1982;62:23–43.
[46] Baltes M, Schneider R, Reuss M. Optimal experimental design for parameter estimation in unstructured growth models. Biotechnolgoy Prog 1994;10:480–488.
[47] Powell MJD. On search directions for minimization algorithms. Math Program 4 1973;4:193–201.
[48] Kirchner JW. Uncertainty Analysis and Error Propagation. University of California, Berkeley: 2001.

Paper appendix

A.1. Example of model application

Compared to the existing models, the one developed in this study can provide information about the uncertainty of the property prediction. This can be considered as of high importance for the use in preliminary risk assessment.

In order to exemplify the simplicity of the model application, the prediction of the heat of combustion ΔH_c° including uncertainty is shown by the example of Methylcyclohexane.

1) The MG GC parameters of the compound have to be identified according to the rules set by Marrero and Gani[12] (Figure 1 and Table 1). The structure of Methylcyclohexane is relatively simple, since it does not contain 3rd order groups. The universal constant from robust regression is $\Delta H_{c const}^\circ$=-99.67 kJ/mol.

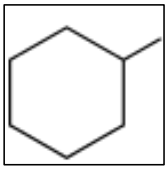


Figure S1. Structure of Methylcyclohexane.

Table S1. Group contribution factors of Methylcyclohexane, obtained from data sheet.

1st order groups	No. N_j	Contribution C_j
CH3	1	-663.99
CH2	5	-607.99
CH	1	-549.79
2nd order group	**No. M_k**	**Contribution D_k**
CHcyc-CH3	1	-2.11

2) The overall equation model equation can be simplified. ΔH_c° can be calculated according to Eq. (29). The unit of ΔH_c° is *kJ/mol*.

$$\Delta H_c^\circ = \Delta H_{c\,const}^\circ + \sum_j N_j C_j + \sum_k M_k D_k + \sum_l E_l O_l \tag{28}$$

$$\Delta H_c^\circ = \Delta H_{c\,const}^\circ + \sum_j N_j C_j + \sum_k M_k D_k = -4'355.51\ kJ/mol \tag{29}$$

3) Using the local parameter covariance matrix $COV(\theta^*)$ and the local sensitivity matrix *J* (see Table 2), obtained from the data sheet, for the respective groups, it is possible to compute the respective confidence interval for the prediction as depicted in Eq. (30) and (31).

Table S2: Parameter covariance matrix $COV(\theta^*)$ and local sensitivity matrix *J*.

	$\Delta H_{c\,const}^\circ$	CH3	CH2	CH	CHcyc-CH3
$\Delta H_{c\,const}^\circ$	294.66				
CH3	-146.81	78.63			
CH2	-0.07	-0.55	0.18		
CH	146.48	-83.25	0.82	114.32	
CHcyc-CH3	0.23	-0.32	0.14	-0.84	79.59
$\frac{\delta\Delta H_c^\circ}{\delta\Delta H_{c\,const}^\circ}$	$\frac{\delta\Delta H_c^\circ}{\delta CH3}$	$\frac{\delta\Delta H_c^\circ}{\delta CH2}$	$\frac{\delta\Delta H_c^\circ}{\delta CH}$	$\frac{\delta\Delta H_c^\circ}{\delta CHcyc-CH3}$	
1	1.31	1.84	0.16	0.05	

$$\delta\Delta H_{c\,1-\alpha_t/2}^{\circ pred} = \delta\Delta H_c^{\circ pred} \pm \sqrt{diag(J(\theta^*)COV(\theta^*)J(\theta^*)^T)} \cdot t(n-p, \alpha_t/2) \tag{30}$$

$$\delta\Delta H_{c\,1-\alpha_{t/2}}^{\circ pred} = 4'355.53\ kJ/mol \pm 15.09\ kJ/mol \tag{31}$$

A.2. Data sheets

The supplementary excel file contain the MG GC factors and the universal constant from robust regression fit without outliers. Furthermore, the parameter correlation matrix and the jacobian are provided. The data sheets are permanently available on the webpage of the Journal of Chemical and Engineering Data: http://dx.doi.org/10.1021/acs.jced.5b00750.

Paper B

Outlier treatment for improving parameter estimation of group contribution based models for upper flammability limit

Jérôme Frutiger, Jens Abildskov, Gürkan Sin

Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

published in Proceedings of the 25th European Symposium on Computer Aided Process Engineering *ESCAPE 25* (ISBN: 978-0-444-63429-0), 37, pages: 503-508, 2015,
DOI: http://dx.doi.org/10.1016/b978-0-444-63578-5.50079-7

and presented at *ESCAPE* 25, May 31 - June 4 2015, Copenhagen, Denmark.

Abstract

Flammability data is needed to assess the risk of fire and explosions. This study presents a new group contribution (GC) model to predict the upper flammability limit UFL of organic chemicals. Furthermore, it provides a systematic method for outlier treatment in order to improve the parameter estimation of the GC model. The new method identifies and removes outliers based on the empirical cumulative distribution plot. It is compared to outlier detection based on Cook's distance and normal cumulative distribution.

1. Introduction

Flammability data provide important information in order to quantify the risk of fire and explosion in process safety studies and assessments. The upper flammability limit (UFL) is defined as the highest possible concentration of a substance in air at which a flammable mixture is formed (Crowl and Louvar, 2013). Experimental data on UFL are not always available due to cost and time constraints in particular at the early stage of process development where alternative concepts are evaluated and ranked before proceeding for more detailed analysis. Property prediction models can in this case be used to estimate the desired flammability data.

Group contribution (GC) based prediction methods use structurally dependent parameters in order to determine the property of pure components. The aim of the parameter estimation is to find the best possible set of model parameters that fits the experimental data.

However, outliers from the model set can strongly influence the parameter estimation, such that the property prediction can be inaccurate in the end. Therefore, it is necessary to identify possible outliers and remove them from the experimental data set. The outlier detection should be simple, following the structure of the model and mathematically consistent.

The Cook's distance (Cook, 1977) measures, how large the degree of influence of a data point is on the parameter set. A large Cook's distance indicates an outlier.

The approach suggested by Furguson, (1961) considers residuals between the experimental and the predicted values as observations coming from a hypothetical distribution, i.e. the normal distribution. The latter can be used in order to determine the outliers of the model.

In this study a methodology for outlier identification based on empirical cumulative distribution function (empirical CDF) is suggested and compared to the outlier treatments using normal probability and Cook's distance.

2. Methodology

2.1. Property model structure according to Marrero and Gani group contribution method

The Marrero Gani (MG) method considers the group contribution in three levels: The contributions from a specific functional group (1st order parameters), from polyfunctional (2nd order parameters) as well as from structural groups (3rd order parameters). Eq. (1) shows its general form.

$$f_i(\mathrm{X}) = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l E_l O_l \qquad f(\mathrm{X}) = T \cdot \theta \tag{1}$$

In Eq. (1) C_j is the contribution of the 1st order group of type-*j* that occurs N_j times whereas D_k is the contribution of the 2nd order group of type-*k* that occurs M_k times in the molecular structure of a pure component. E_l is the contribution of the 3rd order group of type l that has O_l occurrences. The function $f(X)$ is specific for a certain property X (Marrero and Gani, 2001). The parameters can be summarized into the vector θ and the occurrences of the groups can be depicted in the matrix T. As examples, the different GC-factors of Adiponitrile and Methacrylonitrile are visualized in figure 1.

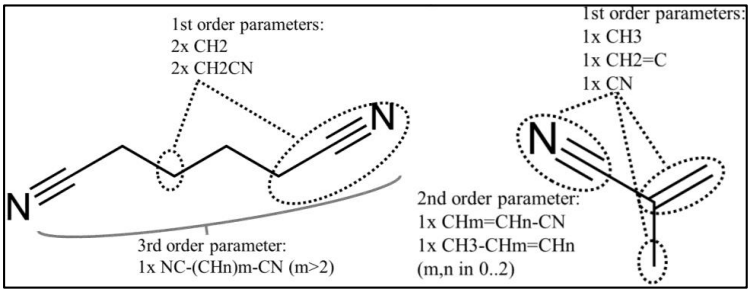


Figure 1. Example of GC-factors of Adiponitrile (left) and Methacrylonitrile (right).

The underlying assumption of group contribution principles is that a certain function of the property is linearly dependent on the contributions of the functional groups. Plots of various classes of pure components versus their increasing carbon number suggest that the appropriate form of the model function is logarithmic as specified in Eq. (2), where UFL_{const} is a universal constant. A selection of classes of compounds is showed in figure 2.

$$f_i(UFL) = \log\left(\frac{UFL_i}{UFL_{const}}\right) \tag{2}$$

2.2. Flammability limit data

Experimental data for UFL of 371 compounds, which includes alkanes, alkenes, alkynes, alcohols, aldehydes, halogenated substances, esters, aromatics and nitrogen compounds, are used from AIChE DIPPR 801 Database (AIChE, 2014). Flammability limits of each compound are provided in percentage volume in air at 298 K.

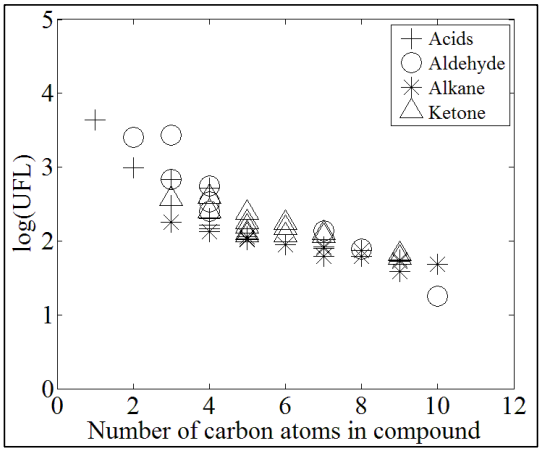


Figure 2. UFL of a selection of pure components versus their increasing carbon number.

2.3. Initialization and sequential parameter estimation

A first guess for the a priori unknown universal constants and MC-GC factors is provided by an approximation using linear algebra according to equation (3).

$$\hat{\theta} = (T^T T)^{-1} \cdot T^T \cdot f(UFL) \tag{3}$$

A value for the constant UFL_{const} is assumed in order to calculate the first guess for the parameters $\hat{\theta}$ from UFL data and the occurrence matrix *T*.

Afterwards the universal constants as well as the 1st, 2nd and 3rd order parameters are estimated separately applying ordinary non-linear regression according to Eq. (4).

$$\theta^* = \arg\min \sum_i (y_i^{\exp} - y_i^{pred})^2 \tag{4}$$

where θ^* corresponds to the estimated parameters and y_i^{exp} to the measurement and y_i^{pred} to the predicted value of compound *i*.

2.4. Simultaneous parameter estimation

The result of the sequential estimation serves as initial guess for the simultaneous parameter estimation algorithm, where all the parameters are estimated together for the ordinary non-linear regression problem.

2.5. Outlier treatment

Three ways of outlier treatment are applied separately to identify the corresponding outliers. The latter are then excluded and the parameter estimation is performed again. The three outlier treatments are compared and evaluated.

Outlier treatment using Cook's distance

Cook's distance measures the influence of a particular data point on the parameter estimation. A particular Cook´s distance D_i can be assigned to every value *i* according to Eq. (5) (Cook, 1977). By definition compounds having a Cook´s distance larger than 4 divided by the number of data *n* is considered as an outlier (Hardin and Hilbe, 2007).

$$D_i = \frac{\sum_{k=1}^{n} (y_k^{pred} - y_{k(i)}^{pred})^2}{p \cdot MSE} \qquad D_i > \frac{4}{n} \tag{5}$$

In Eq.(5) y_k^{pred} is the prediction from the full regression for compound *k*, whereas $y_{k(i)}^{pred}$ is the prediction for compound *k* from a refitted regression where observation *i* is excluded. *MSE* is the mean square error and *p* is the number of parameters.

Outlier treatment using normal distribution

The underlying assumption is that the residuals between the predicted and experimental values are normally distributed and centred around zero (Ferguson, 1961). 95% of the data lie within two times the standard deviation σ of the residuals $y_j^{pred} - y_j^{pred}$. Residuals that fall outside this range of values that could reasonably be expected to occur, i.e. plus and minus 2 times the standard deviation σ, can be considered as outliers. The criterion is formulated in Eq. (6) for the data point *j* to be an outlier.

$$r_j = \left| y_j^{pred} - y_j^{pred} \right| > 2\sigma \tag{6}$$

Outlier treatment using empirical cumulative distribution

In this study we suggest a methodology where the residuals are not assumed to be normally distributed. The empirical cumulative distribution function (empirical CDF) tries instead to estimate the true underlying CDF. For the current data the empirical is depicted in figure 3.

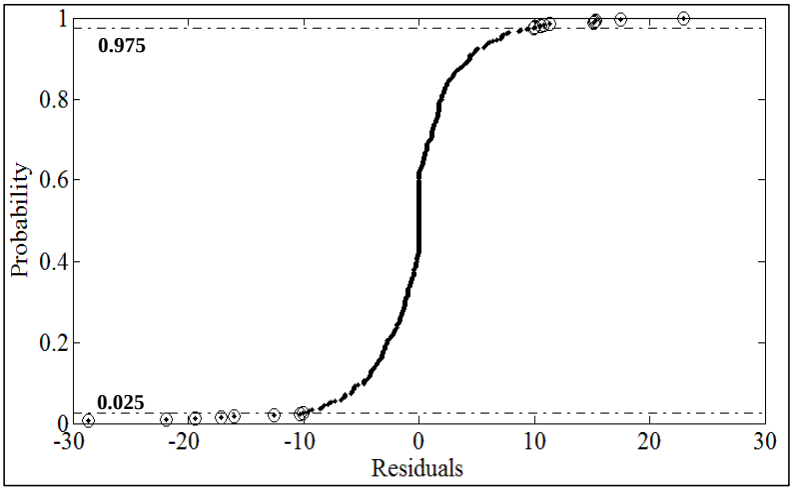


Figure 3. Empirical cumulative distribution of residuals. The data points outside below 0.025 and above the 0. 975 probability levels are considered as outliers.

The empirical CDF is a step function that increases by *1/n* in every data point.

Data points that are not reasonably likely expected to occur according to the empirical CDF can be considered as outliers, i.e. data points that lie below the 2.5% or above the 97.5% probability levels (see figure 3).

3. Results and discussion

The performance statistics of the regression before and after outlier exclusion is depicted in table 1. R^2 is the squared Pearson correlation coefficient between the experimental and the predicted values, N_{out} is the number of outliers removed, *SD* is the standard deviation, *ARE* is the average relative error between the predicted and the experimental values, *SSE* is the sum of squared errors, P_{rc} represents the percentage of the experimental data-points found within ± 25% relative error range respectively [7]. In figure 4 the identified outliers are compared by depicting the prediction versus the experimental value. The equality line indicates a perfect fit.

Table 1. Model performance statistics

	R^2	N_{out}	*SD*	*AAD*	*ARE*	*SSE*	P_{rc} *25%*
Before outlier exclusion	0.72	0	6.2	3.0	24	14,398	66
After exclusion of Cook´s outliers	0.81	18	3	1.7	16	3,106	77
After exclusion of normal probability outliers	0.93	18	2.5	1.7	16	2,284	78
After exclusion of empirical CDF outliers	0.93	20	2.5	1.7	16	2,247	78

All three outlier treatments improve the regression performance. The outlier treatment using empirical cumulative distribution improves the regression similar to outlier detection using normal probability.

Considering figure 4, there are large value data points that are identified as outliers according to Cook's distance even though these compounds match the model. They influence the regression strongly and are assigned a large Cook's distance. Hence, large data points are removed from the data set, even though their prediction matches the experimental value and therefore should not be deemed as outliers. Although removing Cook's distance outliers improves the performance statistics, this outlier treatment is not recommended, because it also removes data points which are clearly in perfect agreement with the model prediction trends.

However, for the outlier treatment using normal CDF and empirical CDF approach of outlier identification, the potential outliers are those that differ a lot from the equality-line. The advantage of empirical CDF is that it does not a priori assume the property to be normally distributed, but tries to reveal the real underlining CDF. In cases where residuals follow a

normal distribution, the difference between our approach that uses empirical distribution function and the Ferguson's method will indeed be negligible as is the case in this contribution. In that case, the empirical CDF will approximate a normal CDF distribution hence verifying the validity of the Ferguson's approach. However, in cases where residuals do not confirm to normal distribution and in particular the residual distribution displays long tails – as can be observed in several property model fittings in Hukkerikar et al. (2012), the empirical CDF approach is expected to guide better outlier detection hence model fitting. A potential caveat of using empirical CDF is that the number of data points (i.e. the size of sample used to construct the cumulative probability function is expected to be representative of residuals distribution) will influence the outcome hence it is recommended to use the empirical CDF approach for relatively large size of data samples as often the case in building GC models. Table 2 shows an example of the prediction of UFL for Methacrylonitrile, which is depicted in figure 1, using model Eq. (1) and (2).

Table 2. Example of the prediction of UFL for Methacrylonitrile using model Eq. (1).

UFL_{const}		123.29
1st	CH3	-1.11
	CH2=C	0.52
	CN	-0.69
2nd	CH3-CHm=CHn	-0.53
	CHm=CHn-CN	-0.61
3rd	none	-
Predicted	*UFL*	10.96
Experim.	*UFL*	11.00

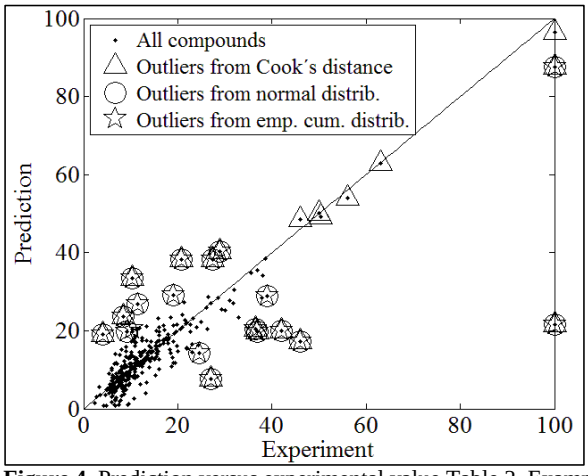


Figure 4. Prediction versus experimental value Table 2. Example of prediction of with identified outliers. UFL for Methacrylonitrile.

4. Conclusion

The study shows a novel prediction model for UFL using the MR-GC method. In this scope an outlier treatment methodology based on the empirical CDF is suggested. The performance statistics of the regression improves similarly to the use of normal CDF. However, the usage of empirical CDF is not restricted by the assumption that residuals should follow a normal distribution and hence have a wider application range. The usage of Cook's distance as an outlier detection method is found unreliable with high number of false detection of data points. This is not surprising since Cook's distance is in fact just a measure of data influence on the regression. High data influence is necessary but not sufficient condition of an outlier since data points in good agreement with regression line may also have a high influence on the regression.

References

[1] Crowl DA, Louvar JF. Chemical process safety. 2nd ed. New Jersey: Prentice Hall International Series in the Physical and Chemical Engineering Sciences; 2013.
[2] Cook DR. Detection of influential observation in linear regression. Technometrics 1977;19:15–8.
[3] Ferguson TS. On the rejection of outliers. Proc Berkeley Symp Math Stat Probab 1961;1:253–87.
[4] Marrero J, Gani R. Group-contribution based estimation of pure component properties. Fluid Phase Equilib 2001;183–184:183–208.
[5] Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE) 2014.
[6] Hardin JW, Hilbe J. Generalized Linear Models and Extensions. 2nd ed. State Press; 2007.
[7] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.

Paper C

Group-contribution based property estimation and uncertainty analysis for flammability-related properties

Jérôme Frutiger, Camille Marcarie, Jens Abildskov, Gürkan Sin

Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

published in Journal of Hazardous Materials (ISSN: 0304-3894), 318, pages: 783–793, 2016.
DOI: http://dx.doi.org/10.1016/j.jhazmat.2016.06.018

Abstract

This study presents new group contribution (GC) models for the prediction of Lower and Upper Flammability Limits (LFL and UFL), Flash Point (FP) and Auto Ignition Temperature (AIT) of organic chemicals applying the Marrero/Gani (MG) method. Advanced methods for parameter estimation using robust regression and outlier treatment have been applied to achieve high accuracy. Furthermore, linear error propagation based on covariance matrix of estimated parameters was performed. Therefore, every estimated property value of the flammability-related properties is reported together with its corresponding 95%-confidence interval of the prediction. Compared to existing models the developed ones have a higher accuracy, are simple to apply and provide uncertainty information on the calculated prediction. The average relative error and correlation coefficient are 11.5% and 0.99 for LFL, 15.9% and 0.91 for UFL, 2.0% and 0.99 for FP as well as 6.4% and 0.76 for AIT. Moreover, the temperature-dependence of LFL property was studied. A compound specific proportionality constant (K^{LFL}) between LFL and temperature is introduced and an MG GC model to estimate K^{LFL} is developed. Overall the ability to predict flammability-related properties including the corresponding uncertainty of the prediction can provide important information for a qualitative and quantitative safety-related risk assessment studies.

1. Introduction

The safety characteristics of hazardous substances provide indispensable information for the risk assessment of chemical products in industrial and domestic processes. In particular flammability-related properties such as the lower and upper flammability limit (LFL and UFL), the flash point (FP) and the auto ignition temperature (AIT) are important to quantify the risk of fire and explosion. In the early design phase a large amount of alternative products and processes are generally analysed, compared and ranked. Whenever experimental values are unavailable property prediction models become a valuable tool [1].

Group contribution (GC) based property models try to estimate a chemical property based on structurally dependent parameters. GC methods are known to be advantageous compared to *ab initio* procedures, quantitative structure property relationship (QSPR) or prediction based on artificial neural networks (ANN), because they are easy to apply, computationally less demanding and have a wide application range [2]. Frutiger et al. [3] stressed the need for thorough parameter estimation and uncertainty analysis for GC models in order to obtain accurate and reliable property predictions. For safety-related properties the provision of uncertainty information (i.e. the upper and lower bound of the 95%-confidence interval) is of particular interest, because the statistical uncertainty should be taken into account, when risk calculations are being carried out [4]. However, there is still a lack of application of uncertainty analysis techniques for safety-related property prediction.

The lower flammability limit (LFL) and the upper flammability limit (UFL) are defined as the lowest and the highest possible concentration of a substance in air at which a flammable mixture is formed. These concentrations are stated at a specific temperature (298K) and pressure (1 atm). However, LFL and UFL change with increasing temperature [5]. The flash point (FP) is the lowest temperature where a liquid forms an ignitable vapour-air mixture. The auto ignition temperature (AIT) is the lowest possible temperature above which a substance will ignite in air without an external ignition source [6].

The review of Vidal et al. [7] provides an overview of the abundant literature, which is available on single point calculations of LFL and FP. Rowley et al. [8] compared extensively a

large variety of the developed methods to estimate LFL at a predefined temperature of 298K (single point prediction). The comparison contains purely correlation-based, GC methods and also detailed mechanistic models. Among the GC based models for LFL and UFL prediction there are several methods suggested in the literature. Shimy [9] derived formulas for different classes of chemicals relating the number of carbon atoms with LFL. Solovev et al. [10] as well as Oehley [11] used atomic indices to calculate LFL. Shebeko et al. [12] used atom and bond connectivity indices in order to model LFL and UFL of pure compounds. Kondo et al. [13][14] developed a GC method to estimate the ratio between LFL and UFL, which they called F-number. All of these methods are simple and easy to apply, but employ very little structural information on the molecules and a limited application range. Hence, the average relative error is high considering different classes of chemicals [8]. Seaton [15] developed a GC method for LFL and UFL of pure compounds. The application range of the latter method is limited by the relatively small number of functional groups. The methods of Shebeko and Seaton have been used to predict non-experimental property values for LFL in the DIPPR 801 database [16]. Albahri [17] developed a structural GC method to predict LFL and LFL. A QSPR model for LFL has been developed by Gharagheizi [18]. Pan et al. [19][20] used topological, charge, and geometric descriptors to describe a QSPR model for LFL and UFL. Recently, Gharagheizi [21] as well as Albahri [22] calculated GC-factors for LFL using artificial neural networks (ANN). Furthermore, Gharagheizi [23] developed a QSPR model for UFL. In a similar approach using ANN, Lazzús [24] predicted the LFL and UFL of various organic compounds. Bagheri et al. [25] used a nonlinear machine learning model to develop a LFL QSPR method. However, the mathematical structure of the latter methods using ANN or machine learning approaches for LFL and UFL is very complex, making model building very tedious. High et al. [26] set up a simple GC model with a limited amount of groups for UFL and included estimations of the upper and lower bound of the confidence limits. Shu et al. [27] presented a method using the threshold temperature (e.g. the ignition temperature) to evaluate UFL of a hydrocarbon diluted within an inert gas. The same authors also presented a model to evaluate the flammable zones of hydrocarbon-air-CO2 mixtures based on flame temperature theory [28]. Rowley et al. [8] provided a GC method that is based on the relationship between LFL, the respective enthalpies of the substance as well as air and the adiabatic flame temperature, obtaining high accuracy. Mendiburu et al. [29][30] developed semi empirical methods for determination of LFL and UFL of C-H compounds, which took into account the stoichiometry of combustion process and the estimation of the adiabatic flame temperature. Except to High et al., none of the above mentioned methods includes a thorough uncertainty analysis. Hence, no information about the respective 95% confidence interval for a specific prediction of LFL and UFL is provided.

The temperature-dependence of LFL and UFL of organic compounds is generally depicted by the modified Burgess-Wheeler law [31], that relates LFL, temperature, the heat capacity of the fuel-air mixture and the heat of combustion ΔH_c°. Britton et al. [32][33] suggested correlations between LFL and the adiabatic flame temperature. Both methods assume that the adiabatic flame temperature is independent of the initial temperature, which was found to be only true for experimental condition, where LFL was measured in a narrow tube [8][34]. A purely empirical correlation of LFL on a wide range of temperature has been proposed by Catoire et al. [35] taking into account the corresponding stoichiometric mixture of fuel and air mixture and the number of carbon atoms in the molecule. However, the model strongly depends on the data set itself. Rowley et al. [8] improved the modified Burgess-Wheeler law by taking into account the temperature-dependence of the adiabatic flame temperature and relating it to the number of carbon atoms. However, there is only limited amount of structural information of the molecules (i.e. the carbon number) taken into account.

Hukkerikar et al. [36] developed a GC model using Marrero/Gani (MG) method for FP and AIT including an uncertainty analysis based on the parameter covariance matrix and

performance criteria to assess the quality of parameter estimation. Frutiger et al. [3] developed a GC model for the heat of combustion ΔH_c° taking into account different parameter regression methods, optimization algorithms, alternative uncertainty analysis methods and advanced outlier treatment. The same authors also analyzed parameter identifiability issues as the source of prediction inaccuracy and uncertainty. Furthermore, they calculated and reported the 95% confidence interval of GC model predictions (prediction accuracy). This thorough and systematic methodology led to significant improvement of GC based model development.

In this study, we therefore aim to provide a new set of improved group contribution models using Marrero/Gani (MG) method [37] to estimate LFL and UFL, FP and AIT at standard conditions using the systematic model development and analysis method of Frutiger et al. [3]. Furthermore, we suggest a GC method to include temperature-dependency in lower flammability limit calculation. The models include a thorough uncertainty analysis (i.e. estimation of the 95%-confidence interval) of every prediction, in order to provide additional information on the reliability of the estimated property. In that sense it is possible to obtain an overall picture of the different flammability properties of a chemical based on the same property prediction methodology.

The paper is organized as follows: (i) the overall methodology for the GC model development and uncertainty analysis for single point LFL, UFL, FP and AIT is shown; (ii) the LFL model is extended to include temperature-dependence; (iii) the performances of the novel GC models are compared with that of existing models; (iv) an application example for 3-Hexanol to calculate LFL including 95% confidence interval is provided.

2. Method

The procedure to develop the GC model for the single point LFL UFL, FP and AIT, to estimate its parameters and to perform the uncertainty analysis, follows the work of Frutiger et al. [3]. Robust regression method as well as the covariance based uncertainty analysis has been applied for this study. Frutiger et al. [3] suggested and compared also alternative methods for parameter estimation and uncertainty analysis, e.g. in order to take into account experimental uncertainties. GC MG factors for FP and AIT are re-estimated using robust regression and outlier treatment, aiming an improved parameter fit compared to the previous estimations [36].

2.1. GC model functions

As a GC model structure the Marrero/Gani (MG) [37] method is chosen, which considers structural contributions on three levels. The MG method is written as

$$f_i(X) = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l \tag{1}$$

$$f(X) = T \cdot \theta \tag{2}$$

A specific functional group (1st order parameters *j*) is expressed by the factor C_j that occurs N_j times. D_k is the contribution factor of the polyfunctional (2nd order parameters *k*) that occurs M_k times in the molecular structure. Finally structural groups (3rd order parameters *l*) are taken into account by the contribution E_l that has O_l occurrences. The function *f(X)* needs to be specified for a certain property *X*. The factors can be determined for a specific molecule following the rules of Marrero et al. [37]. The GC parameters can be summarized in vector θ with *T* being the occurrence matrix of the factors (see Eq. (2)). MG groups are shown for methacrylonitrile and adiponitrile in Figure 1.

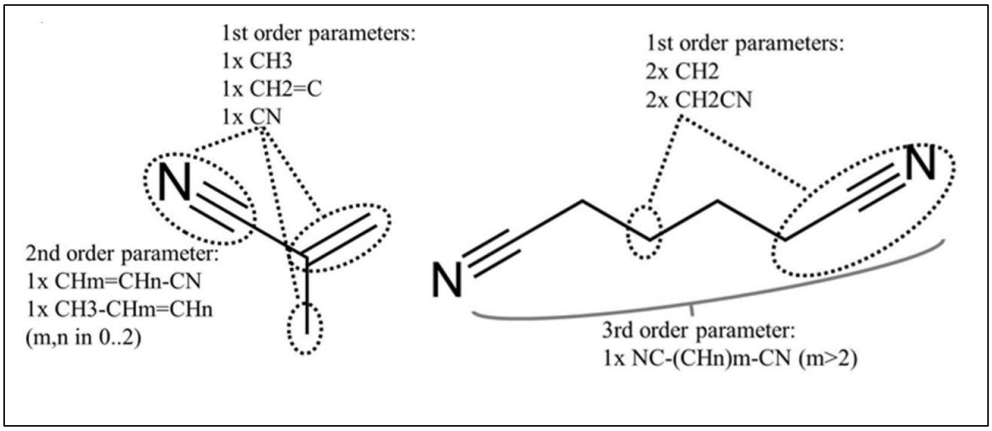


Figure 1. Example of GC-factors of methacrylonitrile (left) and adiponitrile (right).

By plotting various classes of pure components versus their increasing carbon number in homolog series one can obtain ideas regarding the property function *f(X).* Such a homologue series plot is shown for LFL in Figure 2.

The following functions are suggested for LFL, UFL and FP for a specific compound.

$$f(LFL) = log\left(\frac{LFL}{LFL_{const}}\right) = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l \tag{3}$$

$$f(UFL) = log\left(\frac{UFL}{UFL_{const}}\right) = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l \tag{4}$$

$$f(FP) = FP - FP_{const} = \sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l \tag{5}$$

In Eq. (3) to (5) LFL_{const}, UFL_{const}, and FP_{const} are universal constants that need to be determined by the parameter regression. For AIT data on the homolog series suggest a more complex structure involving two summations:

$$AIT = AIT_{const1} + \left(\sum_j P_j X_j + \sum_k Q_k Y_k + \sum_l R_l Z_l\right) \\ . \quad + AIT_{const2} \cdot 10^{-\left(\sum_j N_j C_j + \sum_k M_k D_k + \sum_l O_l E_l\right)} \tag{6}$$

Eq. (5) and Eq. (6) has already been proposed by Hukkerikar et al. [36]. Here more experimental data points are taken into account and a comprehensive methodology for parameter estimation and uncertainty analysis is applied to estimate the GC factors in this study. Eq (4) was first suggested by Frutiger et al. [38], but no thorough parameter estimation and uncertainty analysis has been performed.

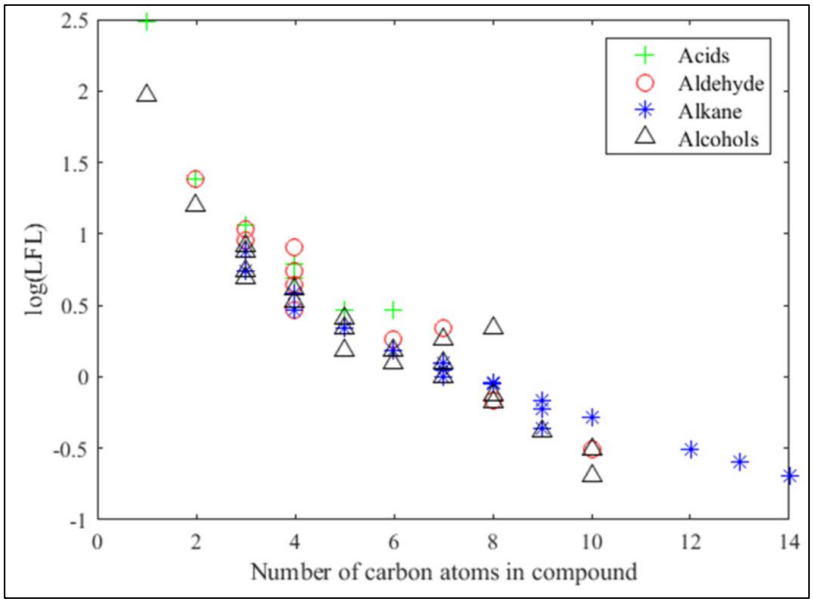


Figure 2. Graphical representation of number of carbon atoms versus property for logarithm of LFL for a selection of groups of pure components.

In order to account for the temperature-dependence of LFL the approach of Rowley et al. [8] is used as a basis to derive a new MG GC method. The latter authors also provided a detailed derivation and explanation of the following equations.

The temperature-dependent LFL of Rowley et al. is based on the following energy balance of the combustion process:

$$LFL(T) \cdot (-\Delta H_c^\circ) + C_{p\,fuel-air} \cdot (T - 298K) = C_{p\,prod} \cdot (T_{ad} - 298K) \tag{7}$$

where ΔH_c° is the heat of combustion, $C_{p\,fuel-air}$ is the heat capacity of the compound and air $C_{p\,prod}$ is the heat capacity of the combustion products and T_{ad} is the adiabatic flame temperature. Rowley et al. further assumed:

1) $C_{p\,fuel-air}$ to be roughly equal to $C_{p\,prod}$ 2) the adiabatic flame temperature T_{ad} as linearly decreasing with increasing initial temperature [34].

This leads to the following generalization of the Burgess-Wheeler law [8]:

$$LFL(T)\cdot = LFL(T_{ref}) + \frac{(1-\gamma)\cdot C_{p\,fuel-air}}{(-\Delta H_c^\circ)} \cdot (T - T_{ref}) \tag{8}$$

where $C_{p\,fuel-air}$ is assumed to be

$$C_{p\,fuel-air} = LFL(T_{ref}) \cdot C_{p\,fuel} + (100\% - LFL(T_{ref})) \cdot C_{p\,air} \tag{9}$$

γ is the compound specific linear constant of T_{ad}, $C_{p\,fuel}$ is the heat capacity of a specific compound at the reference temperature T_{ref} and $C_{p\,air}$ is the heat capacity of air at the reference temperature T_{ref}.

Comparing experimental flammability data for different temperatures and various compounds, usually a linear dependence between LFL and the temperature T is reported by [5][34][39]. Based on this premise, we present a simplified model as follows:

$$LFL_i(T)\cdot = LFL_i(T_{ref}) + K_i^{LFL}\cdot(T - T_{ref}) \tag{10}$$

where K_i^{LFL} is the proportionality constant between LFL and T for a specific compound *i*. K_i^{LFL} could be determined for a certain compound *i* by analyzing the experimental work of Coward et al. [39] and Rowley et al. [34]. Plotting K^{LFL} versus the corresponding carbon number of the compounds implies the possibility of describing this constant by GC models using a reciprocal model function (see Figure 3). Therefore, we propose the following Marrero/Gani GC model to estimate K^{LFL} for a specific compound:

$$f(K^{LFL}) = \frac{K_{const}^{LFL}}{K^{LFL}} = \sum_j N_j C_j \tag{11}$$

with K_{const}^{LFL} as the universal correlation constant and C_j the first order parameters that occurs N_j times.

Comparison with the generalized Burgess-Wheeler law in Eq. (8) with Eq. (10), shows that our proposed proportionality constant can be considered as a lumped parameter of several properties:

$$K^{LFL} = \frac{(1-\gamma)\cdot C_{p\,fuel-air}}{(-\Delta H_c^\circ)} \tag{12}$$

Calculating K^{LFL} directly from GC factors reduces the amount of parameters in the model which makes it easier to apply. Furthermore, it lumps properties that showed to be correlated with increasing carbon number or structurally-dependent group contribution factors in previous studies: $C_{p\,fuel-air}$ is linearly depending on the heat capacities $C_{p\,fuel}$ and $C_{p\,air}$. Joback and Reid depicted the dependence of the heat capacity on the structurally dependent parameters [40]. ΔH_c° is strongly depending on the carbon numbers and a MG GC method has been developed by Frutiger et al. [3]. Rowley et al. [8] showed dependence of γ on the carbon numbers. If for a compound, no experimental value for the mentioned properties exist, then GC models would be used in order to estimate $C_{p\,fuel-air}$, ΔH_c° and γ by Eq. (8). In that sense the introduction of K^{LFL} summarizes structural dependence on the temperature-dependent flammability for a specific compound in one single parameter and provides one single model GC model to estimate it.

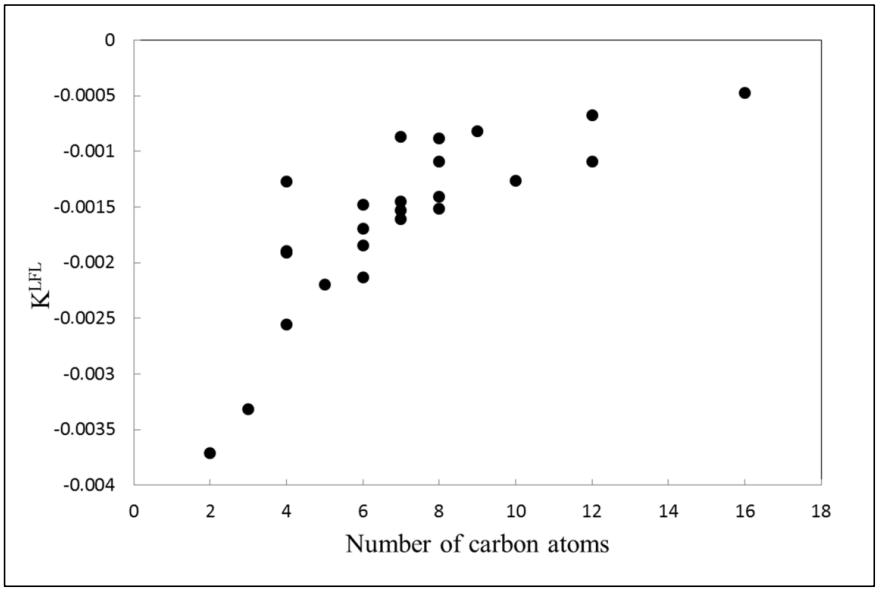


Figure 3. Graphical representation of number of carbon atoms versus K^{LFL}.

2.2. GC parameter estimation and uncertainty analysis

Experimental data for LFL, UFL, FP and AIT are taken from AIChE DIPPR 801 Database [16]. We only considered data points that are classified by DIPPR as "experimental" and "accepted" values. Table 1 shows the number of experimental data points. Data for the temperature-dependence of LFL have been collected from different sources [34][5][39].

Table 1. Number of compounds per property.

	No. compounds
LFL	443
UFL	351
FP	927
AIT	513
K^{LFL}	23

In order to estimate the GC parameters robust regression is chosen, in which the residuals are assigned a certain weight factor w_i, decreasing the influence of experimental data points giving large residuals (not following the model), i.e. potential outliers [41]. Robust regression has been shown to be advantageous over standard non-linear regression for the estimation of GC factors [3].

$$y = \{LFL, UFL, FP, AiT, K_i^{LFL}\} \tag{13}$$

$$\theta^* = arg\ min \quad \sum_i w_i \cdot (y_i^{exp} - y_i^{pred})^2 \tag{14}$$

$$w_i = \frac{1}{1 + (y_i^{exp} - y_i^{pred})^2} \tag{15}$$

θ^* is the parameter (1st, 2nd and 3rd order group contributions) estimates and y_i^{pred} is the prediction of compound *i* according to Eq. (3) to (6) and y_i^{exp} its corresponding experimental value.

Outliers are identified using the empirical cumulative distribution function (CDF) of the residuals between experimental and predicted values, which has been described for GC models by Frutiger et al.[38]. The empirical CDF is defined as a step function increasing by 1/n in every data point. The major advantage of this methodology is that the distribution of the residuals is estimated from the data themselves, not a priori assuming normal distribution. Outliers are considered as data points that that lie below the 2.5% or above the 97.5% probability levels.

The Uncertainty analysis is based on linear error propagation using parameter covariance matrix [3][36].

The covariance matrix, $COV(\theta^*)$ of parameter estimators is asymptotically estimated as follows

$$COV(\theta^*) = \frac{SSE}{n-p}(J(\theta^*)^T J(\theta^*))^{-1} \tag{16}$$

where *p* is the number of parameters, *SSE* is the minimum sum of squared errors given by the regression model, *n* is the number of data points and, *J* is the Jacobian of the model function *f* with respect to the parameter values θ^*. Linear error propagation allows estimating the uncertainty of the property predictions. The covariance matrix of the predictions $COV(y^{pred})$ can be approximated using the Jacobian and the covariance of the parameter estimates as shown in Eq. (16),

$$COV(y^{pred}) = J(\theta^*)COV(\theta^*)J(\theta^*)^T \tag{17}$$

A student *t*-distribution $t(n-p, \alpha_t/2)$ (with $\alpha_t/2$ percentile) can be used to calculate the confidence intervals of the property predictions

$$y_{1-\alpha}^{pred} = y^{pred} \pm \sqrt{diag(J(\theta^*)COV(\theta^*)J(\theta^*)^T)} \cdot t(n-p, \alpha_t/2) \tag{18}$$

where $diag(COV(\theta^*))$ are the diagonal elements of $COV(\theta^*)$ and $diag(J(\theta^*)COV(\theta^*)J(\theta^*)^T))$ the diagonal elements of $COV(J(\theta^*)COV(\theta^*)J(\theta^*)^T))$.

In order to quantify and compare the performance of the parameter estimates the following statistics are calculated: the Pearson correlation coefficient R^2 and the average relative error *ARE*:

$$R^2 = \frac{\sum_j (y_j^{exp} - y_j^{pred})^2}{\sum_j (y_j^{exp} - \mu_y)^2} \tag{19}$$

$$ARE = \frac{1}{n}\sum_{j}\frac{(y_j^{exp}-y_j^{pred})}{y_j^{exp}} \quad (20)$$

with y_j^{pred} the prediction of compound *j*, y_i^{exp} the experimental value and μ_y the mean value.

In order to compare the newly developed temperature-dependent GC model for LFL with the model developed by Rowley et al. [8] Akaike information criterion (*AIC*) [42] is used. AIC is a way of model selection based on information theory, which tries to account for both the goodness of the model fitting and the complexity of the model. Akaike information criterion (*AIC*) is given by Eq. (21).

$$AIC = -2 \cdot log(\frac{SSE}{n}) + 2p \quad (21)$$

SSE is the sum of squared errors, *n* the number of data points and *p* the number of parameters [42].

3. Results and Discussion

3.1. Results of the GC parameter estimation and uncertainty analysis

The results of the parameter estimation using robust regression are shown in Table 2 and Table 3. R^2 is the Pearson correlation coefficient, *ARE* is the average relative error, *SSE* is the sum of squared errors between the experimental and predicted property values and *SD* is the standard deviation. $P_{rc}25$ represents the percentage of the experimental data points found within ± 25% relative error range respectively. The performance statistics show that the GC parameter fits for LFL, UFL and K_i^{LFL} are very good. For FP and AIT the performance statistics of the re-estimated parameters can be compared to the estimation of Hukkerikar et al. [36], who used a standard non-linear regression. As it can be seen in Table 3, robust regression and systematic outlier removal gives a much better parameter fit. Figure 4 depicts the prediction based on the model of versus the experimental values used for the GC parameter estimation of LFL, UFL, FP and AIT. The GC factors of all of the developed models can be found in the supplementary material.

Table 2. Regression model performance statistics for LFL, UFL and γ.

	R^2	*ARE [%]*	*SD*	*SSE*	$P_{rc}25$ *[%]*
LFL	0.99	11.5	0.24	23	88
UFL	0.91	15.9	2.74	77	82
K^{LFL}	0.89	14.7	$6.01 \cdot 10^{-4}$	$7.7 \cdot 10^{-6}$	76

Table 3. Regression model performance statistics for FP and AIT.

	R^2	*ARE [%]*	*SD*	*SSE*	*Prc25 [%]*
FP	0.99	2.0	9.99	$4.73 \cdot 10^5$	100
FP (Hukkerikar et al.)	0.80	3.2	14.30	-	98
AIT	0.76	6.4	41.35	$4.48 \cdot 10^6$	97
AIT (Hukkerikar et al.)	0.72	6.8	56.74	-	96

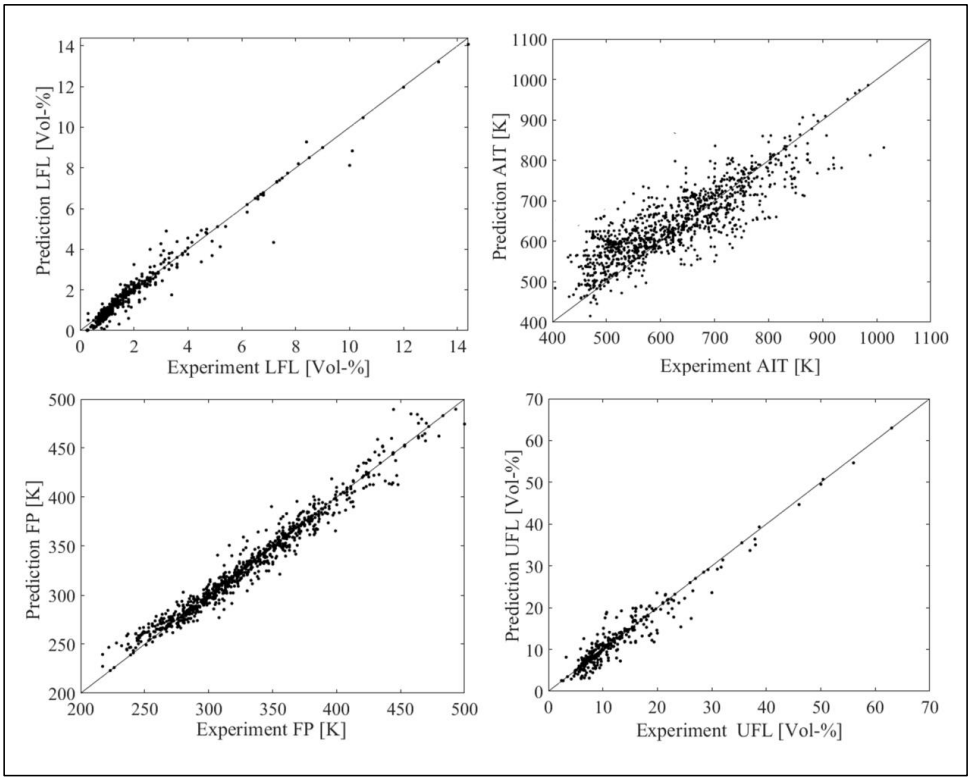


Figure 4. Prediction versus experimental value for LFL, UFL, FP and AIT.

The average relative error *ARE* and the number of data included in this study for the LFL and UFL model are compared to other property prediction models in Table 4 and Table 5.

Table 4. Comparison of developed LFL model with existing GC models. Abbreviations: average relative error (*ARE*), Marrero/Gani (MG), group contribution (GC), atom and bond connectivity (AC), quantitative structure property relationship (QSPR), artificial neural networks (ANN).

	Model structure	*ARE [%]*	*No. of data*
Current study	**MG GC**	**12**	**443**
Oehley, 1953 [11]	AC	27	-
Solovev et al., 1960 [10]	GC	25	-
Shimy, 1970 [9]	CN	24	9
Shebeko et al., 1983 [12]	AC	21	70
Seaton, 1991 [15]	GC	16	152
Kondo et al., 2001 [13]	GC	24	238
Albahri, 2003 [43]	structural GC	10	109
Gharagheizi, 2008 [18]	QSPR	8	1056*
Pan et al., 2009 [19]	QSPR	5	1038*
Gharagheizi, 2009 [23]	ANN	4	1056*
Lazzús, 2011 [24]	ANN	9	328
Rowley et al., 2011 [8]	GC	11	509
Bagheri et al., 2012 [25]	QSPR	1	1615*
Mendiburu et al., 2015 [29]	semi empirical	9	120

*included experimental and predicted property values hence it is not an objective performance evaluation of a model.

Table 5. Comparison of developed UFL model with existing GC models. Abbreviations: average relative error (*ARE*), Marrero/Gani (MG), group contribution (GC), atom and bond connectivity (AC), quantitative structure property relationship (QSPR), artificial neural networks (ANN).

	Model structure	*ARE [%]*	*No. of data*
Current study	**MG GC**	**16**	**351**
Shebeko et al., 1983 [12]	AC	25	70
High et al., 1987 [26]	GC	26	181
Seaton, 1991 [15]	GC	20	152
Albahri, 2003 [17]	structural GC	12	109
Pan et al., 2009 [20]	QSPR	19	588*
Gharagheizi, 2009 [23]	QSPR	10	1057*
Lazzús, 2011 [24]	ANN	7	328
Mendiburu et al., 2016 [30]	semi empirical	8	115

*included experimental and predicted property values hence it is biased.

The comparison could only be made according to the average relative error *ARE*, due to the fact that no uncertainty analysis has been performed by the other authors. The current model provides for every predicted value the corresponding uncertainty, which is lacking in the other models (with the exception of High et al. [26]).

Considering the *ARE* of LFL, the model developed in this study performs better than the previous LFL models of Oehley, Solovev et al., Shimy, Shebeko et al., Seaton and Kondo et al.. Furthermore, the amount of data that are taken into account is much higher for the present model. This increases the application range of the model, since more chemicals from different classes of molecules have been used in the model development. The current LFL model performs similar in comparison to the recent GC prediction method of Rowley et al. and the best performing model of Albahri. The work of Mendiburu et al. took only C-H compounds into account and can therefore not be compared directly to the model of this study. The ANN methods of Lazzús and Albahri shows better performance statistics as well. However, these authors took a lower amount of experimental data points into account for the fitting of their model. Hence, the application range is narrower. Furthermore, the ANN structure is very complex for even a relatively small number of fitted data. In that sense its applicability is more difficult and its application range is smaller. Similar conclusions can be made for UFL, where the developed model is superior to Shebeko, High et al., Seaton and Pan et al.. Albahri and Lazzús perform slightly better, but they used a smaller amount of data points, which leads to a smaller application range.

The ANN and QSPR models of Gharagheizi, Pan et al. and Bagheri et al. for LFL and UFL have a lower *ARE* and more data points. However, the amount of data consist of all experimental data and predicted values available in the DIPPR database which is not a scientifically accepted way to compare model performance statistics. A parameter estimation should solely be based on experimental data points only [44]. While comparing ANN or QSPR

with GC models for flammability, it is important to state that ANN/QSPR and are fundamentally different to GC methods in the sense that the aim is to build the best possible model structure (i.e. considering variables and descriptors). However, the model structure is fixed in GC methods and its goal is to estimate the parameters in the best possible way given a certain available set of experimental data. The structure of the MG GC model is much simpler compared to ANN and easier to apply in practice. Furthermore, whereas the reliability of the GC model predictions have been statistically demonstrated and verified against application in practice, establishing the reliability and confidence of parameter estimation in ANN or QSPR remains to be demonstrated. Furthermore, GC models allow adding new experimental values to the parameter estimation without changing the model structure. In QSPR and ANN model building need to be performed all over again [3].

GC factors for the LFL, UFL, AIT, and K_{LFL} of a selection of functional groups are depicted in Table 6. The complete list of the GC factors can be found in the supplementary material.

Table 6. Selection of commonly used GC factors for the LFL, UFL, FP, AIT and K_{LFL} model. The complete list of all GC factors can be found in the supplementary material.

		LFL	UFL	FP [K]	AIT [K]		K_{LFL}
		[Vol-%]	[Vol-%]		*Factor (linear)*	*Factor (10^)*	
Type	Constant	4.53	129.96	195.22	561.19	55.19	-0.0036
1st	CH3	-0.24	-1.15	8.32	-74.66	-0.38	1.87
1st	CH2	-0.23	-0.14	12.49	2.19	0.14	-0.20
1st	CH	-0.23	0.89	7.18	94.93	0.61	-1.09
1st	CH2=CH	-0.49	-0.68	18.47	-98.80	-0.31	-0.58
1st	aromaticCH	-0.22	-0.46	13.19	-9.84	-0.13	0.40
1st	aromaticC	0.05	0.20	18.25	-46.00	-34.79	-0.58
1st	OH	0.06	-0.76	69.04	16.20	-0.19	-0.88
1st	COOH	0.00	-1.03	118.40	98.45	-0.03	-1.01
1st	aromaticC-CO	-0.94	0.25	83.76	302.15	8.48	2.14
1st	aromaticC-CHO	-0.07	-0.40	71.33	-46.26	6.07	-1.62
1st	CHNH2	-0.33	-0.19	30.68	235.16	50.42	-
1st	NH2	0.02	-0.13	58.96	-38.58	-0.04	-
1st	-Br	1.00	-1.23	47.63	-94.63	-0.48	-
1st	-F	1.15	-0.56	-9.22	-221.62	-0.69	-
1st	-Cl	0.80	-1.20	21.42	-143.08	-0.58	-
1st	Si	-2.08	3.34	12.06	27.78	0.02	-

2nd	CO-O-CO	0.07	-0.10	5.35	-57.25	-0.25	-
2nd	aromaticC-CH(CH3)2	-0.24	-0.27	4.54	29.19	0.23	-
2nd	aromaticC-C(CH3)3	0.01	0.30	13.22	21.31	-0.71	-
2nd	(CHn=C)(cyclic)-CHO (n in 0..2)	-0.14	-0.10	16.51	-15.50	0.00	-
2nd	(CHn=C)cyclic-CH2 (n in 0..2)	-0.25	0.08	-5.14	-41.08	-0.06	-
2nd	CHcyclic-CH3	0.00	-1.72	1.64	1.43	-0.34	-
2nd	CHcyclic-CH2	-0.02	-2.25	4.37	14.03	2.23	-
2nd	>Ncyclic-CH3	0.01	0.05	-23.86	60.15	0.17	-
3rd	aromaticRINGs1s2	0.12	-0.01	-15.44	134.03	0.22	-
3rd	aromaticRINGs1s3	-0.01	-0.07	-6.41	122.77	0.21	-
3rd	PYRIDINEs2	-0.15	-0.32	-8.10	54.16	0.04	-
3rd	aromatic.FUSED[2]	0.02	0.12	13.03	-16.38	69.66	-
3rd	aromatic.FUSED[2]s1	-0.13	0.23	1.01	4.36	34.82	-

Table 7 gives an example of predicted values for a variety of organic compounds. The large variety of compounds from different chemical classes illustrates the wide application range of the developed models. More predicted values for different compounds can be found in the supplementary material.

Table 7. Predicted values including the respective 95% confidence interval for a variety of selected number of chemical compounds. In all cases experimental data (not shown) falls within 95% confidence interval.

	LFL [Vol-%]			UFL [Vol-%]			FP [K]			AIT [K]		
	Pred.	95% conf.int.		Pred.	95% conf. int.		Pred.	95% conf. int.		Pred.	95% conf. int.	
n-Heptane	0.75	0.63	0.87	6.28	5.10	7.46	274.30	271.41	277.18	611.41	598.42	624.40
2-Methylhexane	0.71	0.59	0.82	6.21	4.84	7.57	266.46	263.26	269.65	607.46	591.98	622.95
3,3-Diethylpentane	0.44	0.25	0.62	4.67	2.61	6.73	282.92	275.80	290.04	657.55	619.25	695.84
Cycloheptane	0.98	0.60	1.37	5.96	4.26	7.67	283.90	276.92	290.88	559.04	523.08	595.00
1-Pentene	1.14	0.97	1.31	9.98	8.20	11.76	246.78	242.92	250.63	601.48	583.79	619.17
1-Octene	0.51	0.42	0.61	6.47	4.91	8.03	284.25	280.49	288.01	599.36	582.47	616.24
Benzene	1.20	0.06	2.34	8.00	0.78	15.22	274.35	267.50	281.21	770.73	730.92	810.55
Toluene	1.09	0.64	1.54	7.87	3.78	11.95	293.78	288.77	298.79	765.55	738.31	792.79
Ethylbenzene	1.04	0.51	1.57	7.25	3.70	10.79	305.57	300.91	310.24	720.08	693.02	747.13
o-Xylene	1.11	0.54	1.68	7.64	2.49	12.78	297.75	291.30	304.21	754.03	715.95	792.12
Propanal	3.12	2.65	3.59	20.42	17.21	23.63	260.91	254.59	267.23	536.40	506.65	566.16
Butanal	2.40	2.00	2.80	17.67	14.68	20.66	273.40	267.15	279.64	536.33	506.71	565.94
Acrolein	2.80	2.06	3.54	29.22	22.91	35.53	259.44	247.96	270.91	517.71	448.45	586.96

Benzaldehyde	1.40	0.26	2.54	8.50	1.25	15.75	332.49	320.44	344.54	556.89	499.89	613.90
2-Heptanone	0.97	0.67	1.28	6.97	4.59	9.35	313.93	307.13	320.73	668.05	633.43	702.66
Ethanol	3.33	2.98	3.68	16.63	14.31	18.96	285.07	281.55	288.59	637.56	619.58	655.54
1-Propanol	2.55	2.28	2.83	14.40	12.44	16.36	297.56	294.19	300.92	636.24	618.67	653.81
Isopropanol	2.08	1.43	2.73	11.26	7.21	15.31	283.21	277.34	289.09	646.61	612.04	681.18
1-Butanol	1.96	1.72	2.20	12.46	10.65	14.27	310.05	306.79	313.30	635.02	617.80	652.24
1-Pentanol	1.50	1.28	1.72	10.78	9.02	12.55	322.54	319.35	325.73	633.89	616.95	650.83
n-Butyric acid	2.21	1.77	2.65	10.94	7.65	14.24	346.91	341.55	352.28	634.80	611.79	657.80
n-Pentanoic acid	1.70	1.35	2.04	9.47	6.56	12.38	359.40	354.09	364.71	633.46	610.76	656.17
Methyl tert-butyl ether	1.42	0.91	1.94	6.90	4.90	8.90	253.62	247.77	259.47	630.16	603.54	656.78
Methyl ethyl ether	2.31	2.00	2.61	15.08	13.20	16.95	245.01	240.91	249.11	578.60	558.38	598.82
Divinyl ether	1.70	0.56	2.84	27.00	19.75	34.25	226.15	203.21	249.09	623.23	492.98	753.48
1,4-Dioxane	2.36	1.87	2.84	21.06	16.18	25.93	284.78	275.73	293.83	538.61	496.87	580.36
tert-Butyl ethyl ether	1.15	0.72	1.57	6.32	4.19	8.44	263.03	257.54	268.51	616.04	592.70	639.37
1,2-Dichloroethane	4.93	4.30	5.56	19.11	13.60	24.62	299.20	292.07	306.32	701.28	651.93	750.64
1,2-Dichloropropane	3.57	2.69	4.45	14.50	7.25	21.75	288.60	281.12	296.08	735.08	663.66	806.51
Isopropyl chloride	2.65	1.92	3.38	18.88	13.59	24.18	246.57	238.40	254.73	712.43	635.55	789.30
1-Chloropentane	1.69	1.48	1.90	8.57	6.94	10.20	292.99	288.99	297.00	653.07	627.34	678.80

Dimethylamine	2.80	1.66	3.94	14.40	7.15	21.65	223.15	200.21	246.09	599.55	457.37	741.73
Triethylamine	1.29	0.24	2.33	8.69	2.88	14.50	264.68	256.03	273.34	554.89	482.22	627.56
Pyridine	1.80	0.66	2.94	12.00	4.75	19.25	287.83	277.45	298.21	735.61	662.44	808.78
Aniline	1.37	0.82	1.91	10.25	3.86	16.65	345.81	339.13	352.48	768.19	724.24	812.13
Thiophene	1.29	0.45	2.13	7.96	3.50	12.41	281.11	268.08	294.14	640.15	569.30	710.99
Dimethyl sulfoxide	2.60	1.46	3.74	28.50	21.25	35.75	361.00	338.06	383.94	492.34	350.16	634.52
n,n-Dimethylacet-amide	1.80	0.66	2.94	11.50	4.25	18.75	336.15	313.21	359.09	631.70	489.52	773.88
2-Methoxyethanol	2.20	1.64	2.76	19.60	15.90	23.30	304.56	299.42	309.70	578.47	547.32	609.63
n-Ethylaniline	1.60	0.46	2.74	9.50	2.25	16.75	360.82	344.60	377.04	697.72	643.08	752.35
Vinyltrichlorosilane	3.71	2.82	4.60	50.72	43.55	57.90	290.02	278.46	301.57	628.90	559.44	698.37
Ethylene glycol monopropyl ether	1.30	0.99	1.60	14.68	12.10	17.26	329.54	324.57	334.52	577.20	546.36	608.04
1-Chloro-2,4-dinitrobenzene	2.19	1.13	3.25	22.00	14.75	29.25	462.69	451.75	473.64	677.46	593.67	761.26
Ethyl lactate	1.69	1.01	2.37	11.40	4.15	18.65	331.00	308.06	353.94	718.18	650.12	786.23
2-Ethoxyethylacetate	1.56	1.19	1.93	11.06	8.55	13.57	320.36	314.75	325.97	654.40	619.29	689.51

Figure 5 shows the results of the covariance-based uncertainty analysis, exemplified for the case of LFL. The experimental and the predicted values of LFL with the respective 95%-confidence interval of the prediction highest value and for every substance are shown. The compounds are ordered from lowest to highest given an index number respectively. The 95%-confidence interval is a narrow band that includes the experimental values. The detailed covariance-based uncertainty analysis is another advantage of the developed GC models. Whereas the majority of the other authors define the quality of their model only with *ARE*, we can provide the 95%-confidence interval for every prediction. This additional information, i.e. the reliability of the prediction, can be vital in the context of a quantitative safety-related risk analysis. For example it is possible to use the lower-bound value of the confidence interval in a conservative analysis approach. In fact, the lower bound of the confidence interval for LFL, is approximately 20% of the LFL values. The latter is commonly used as a rule of thumb in quantitative risk analysis (QRA) studies [45].

Although the extension to mixtures lies far beyond the scope of this work, users can calculate the properties of mixtures from the current pure component model by applying simple mixing rules (e.g. le Chatelier's mixing rule for flammability limit [46]).

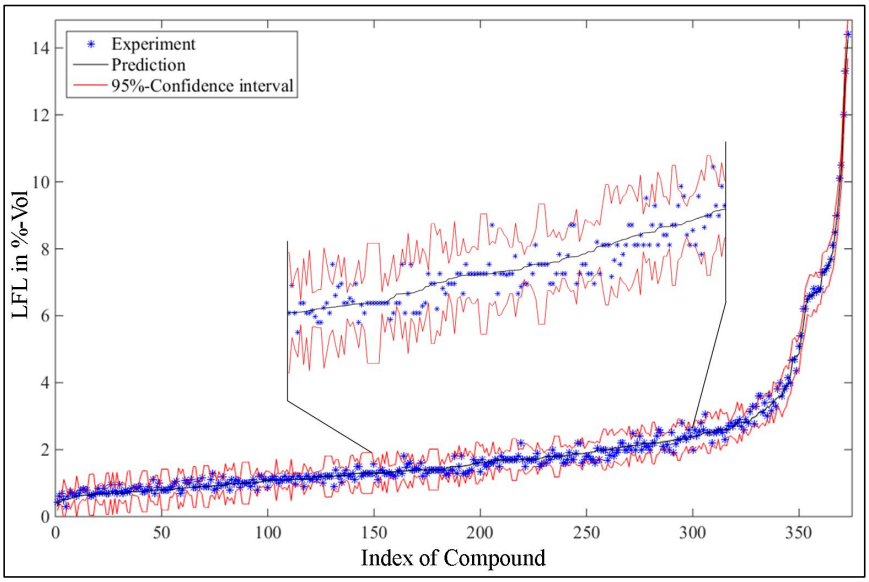


Figure 5. Experimental as well as predicted value of LFL for every compound with 95%-confidence intervals generated by covariance-based uncertainty analysis. A section of the plot is enlarged to show the distribution of the experimental values around the prediction.

The results of the calculation of the Akaika information criterion (*AIC*) for small sample for the developed temperature-dependent model compared to the one developed by Rowley et al. [8] is shown Table 8. The temperature-dependent LFL model developed in this study has been evaluated for different temperatures. These were used to calculate the sum of squared errors and subsequently *AIC*.

Table 8. Akaika information criterion (*AIC*) for small sample with *SSE* (sum of squared errors of the fit), *n* (number of experimental data points), *p* (total number of parameters.

	SSE	*n*	*p*	*AIC*
Current model	3.39	16	22	45.6
Rowley et al.	0.45	16	32	67.3

For the developed model in this study the total number of parameters *p* is consisting of the 21 GC factors, and LFL(T=298K). For the study of Rowley et al. the number of parameters *p* is assumed under assumption that the heat capacity and the heat of combustion needs to be predicted, which is needed if the temperature-dependent LFL is calculated from predicted values only (according to Eq. (8)). The simplest GC based model for the prediction of the heat capacity is Joback and Reid´s method with 20 parameters. The easiest way to predict the heat of combustion is deriving it from the heat of formation using Benson´s method with approximately 12 parameters.

AIC is lower for the newly developed model in this study, although the model of Rowley et al. shows a better fitting to the experimental data. The *AIC* calculation enforces the fact that the newly developed model using the lumped parameter is a sparse model and should therefore be selected.

3.2. Demonstration of model application

The developed models allow calculating the safety-properties from the molecular structure only and include an uncertainty analysis. Figure 6 depicts the result of example calculations with the developed GC MG models for 3-Hexanol. It provides an overall picture of the major flammability property predictions including the corresponding 95% confidence interval.

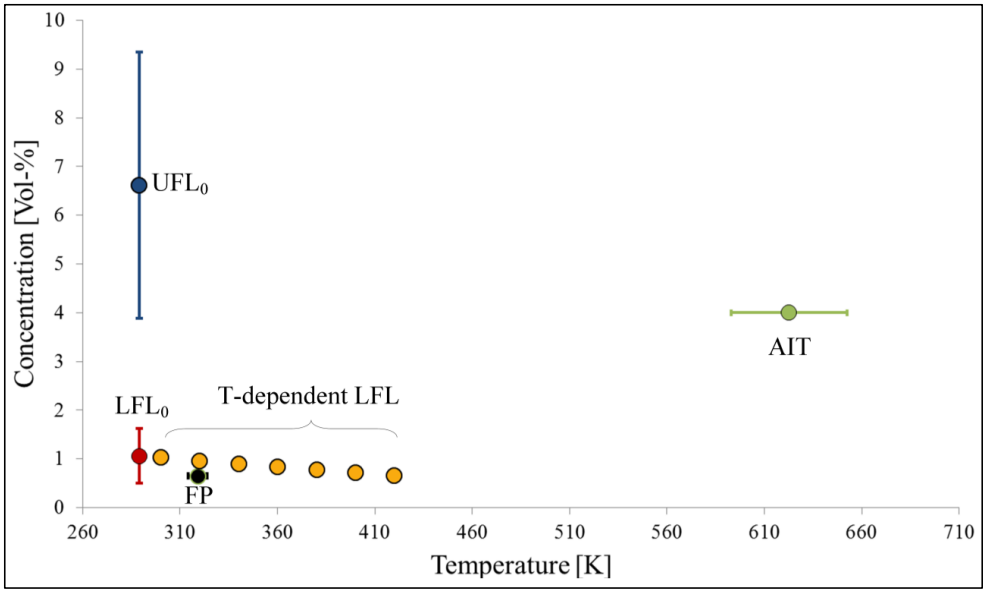


Figure 6. Overview of the generated flammability-related properties by the developed GC MG models including 95% confidence interval: LFL_0 (lower flammability limit at T= 298K), UFL_0 (upper flammability limit T= 298K), FP (flash point), AIT (auto ignition temperature) and Temperature-dependent LFL (without uncertainty).

In order to demonstrate the simplicity of the model application, the prediction of LFL at 298K (single point) including uncertainty and at a different temperature (350K) point and using the temperature-dependence is shown by the example of 3-Hexanol. The calculation procedure for UFL, FP and AIT is analogous. The respective parameter values, covariance matrices and jacobians for the model are given in the supplementary material. Further information (e.g. on the identification of the GC factor for a new molecule) can also be provided by the authors upon request.

1) The MG GC parameters of the compound have to be identified according to the rules set by Marrero and Gani [37]. These rules state how to identify 1st, 2nd and 3rd order parameters from Simplified Molecular Input Line Entry Specification (SMILES). The structure of 3-Hexanol is depicted in Figure 7 and the corresponding Marrero Gani GC factors collected from the supplementary material in Table 9. The structure of 3-Hexanol is relatively simple, hence it does not contain 3rd order groups. The universal constant from robust regression is LFL_{const}=4.53 *Vol-%*.

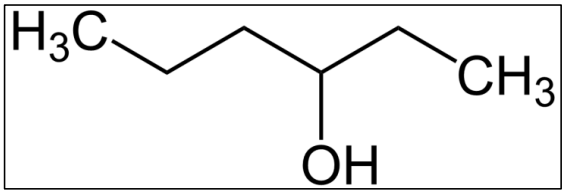


Figure 7. Structure of 3-Hexanol [47].

Table 9. Group contribution factors of 3-Hexanol, obtained from GC factor sheet.

1st order groups *j*	N_j	C_j
CH3	2	-0.24
CH2	3	-0.23
CH	1	-0.23
OH	1	0.06
2nd order group *k*	M_k	D_k
CHOH	1	-0.11

2) The overall model equation according to Eq. (3) can be simplified. The GC factors are taken from Table 9 and can be inserted into Eq. (23). Hence, LFL of 3-Hexanol can be calculated according in Eq. (24). The unit of LFL is Volume-% (*Vol.%*) of the chemical in air.

$$LFL = LFL_{const} * exp(\sum_j N_j C_j + \sum_k M_k D_k + \sum_l E_l O_l) \tag{22}$$

$$LFL = LFL_{const} * exp(\sum_j N_j C_j + \sum_k M_k D_k) \tag{23}$$

$$\begin{aligned} LFL &= 4.53 * exp(2 \cdot -0.24 + 3 \cdot -0.23 + 1 \cdot -0.23 + 1 \cdot -0.06 + 1 \cdot -0.11)\ Vol.\% \\ &= 1.06\ Vol.\% \end{aligned} \tag{24}$$

3) Using the parameter covariance matrix $COV(\theta^*)$ and the sensitivity matrix *J* that can be found in the supplementary material for the respective groups (see Table 10), it is possible to compute the respective confidence interval for the prediction as depicted in Eq. (25) and (26).

Table 10. Parameter covariance matrix $COV(\theta^*)$ and local sensitivity matrix *J*.

	LFL_{const}	CH3	CH2	CH	OH	CHOH
LFL_{const}	0.99					
CH3	-0.11	0.012				
CH2	$-3.1\cdot10^{-4}$	$-1.5\cdot10^{-4}$	$5.1\cdot10^{-4}$			
CH	-0.11	-0.012	$1.6\cdot10^{-4}$	0.025		
OH	-0.093	$-9.6\cdot*10^{-3}$	$-3.4\cdot10^{-5}$	$-9.8\cdot10^{-3}$	0.58	
CHOH	$2.9\cdot10^{-3}$	$-5.3\cdot10^{-5}$	$-3.3\cdot10^{-4}$	$-2.6\cdot10^{-4}$	$-3.4\cdot10^{-3}$	0.042

$\frac{\delta LFL}{\delta LFL_{const}}$	$\frac{\delta LFL}{\delta CH3}$	$\frac{\delta LFL}{\delta CH2}$	$\frac{\delta LFL}{\delta CH}$	$\frac{\delta LFL}{\delta OH}$	$\frac{\delta LFL}{\delta CHOH}$
0.43	1.5	0.97	0.15	$-1.7*10^{-3}$	0.014

$$LFL^{pred}_{1-\alpha_t/2} = LFL_{pred} \pm \sqrt{diag(J(\theta^*)COV(\theta^*)J(\theta^*)^T)} \cdot t(n-p, \alpha_t/2) \quad (25)$$

$$LFL^{pred}_{1-\alpha_{t/2}} = 1.06\ Vol.\% \pm 0.56 Vol.\% \quad (26)$$

The above can be compared to the method of Gmehling et al. [48], who predicted LFL for 3-Hexanol to be 1.29 *Vol.%* [16]. Hence, this value falls squarely within the predicted 95%-confidence interval of the model verifying the reliability of the model for this particular point.

4) In order to account for the temperature-dependency it is necessary to calculate K^{LFL} with the developed MG GC models. LFL of 3-Hexanol should be calculated at a temperature of 320K following Eq. (8) and (9). The value calculated under point 3 can be taken as the reference value $LFL(T_{ref} = 298K) = 1.06\ Vol.\%$.

K^{LFL} according to the developed model in Eq. (11) is given by:

$$K^{LFL}_{3-Hexanol} = \frac{K^{LFL}_{const}}{\sum_j N_j C_j} = 3.05 \cdot 10^{-3} \quad (27)$$

Using the above values, LFL of 3-Hexanol at the specific temperature of 320K is given by:

$$LFL(T = 320) = LFL(T_{ref}) + \cdot K^{LFL}_{3-Hexanol} \cdot (T - T_{ref}) \quad (28)$$

$$LFL(T = 320K) = 0.97\ Vol.\% \quad (29)$$

The uncertainty of the temperature-dependent LFL calculation is huge, around 100% of the estimate value (not shown Eq. (29) and in Figure 6). The reason for this high uncertainty is the limited number of compounds with temperature-dependent LFL data, hence the term $t(n-p,\alpha_t/2)$ becomes very large for K^{LFL}. Higher reliability of the prediction can only be achieved by using more experimental data for the parameter estimation. However, the new GC based temperature-dependent model demonstrated above provides an approximation where the true value of the LFL is most likely to be found. The latter is useful for performing safety analysis under lack of experimental data, which is the second best alternative.

4. Conclusion

In this study, a new GC method has been developed for the calculation of LFL and UFL as well as a new model for estimating temperature dependence of LFL. Furthermore, the parameters for the previous model of FP and AIT have been improved thanks to expanded data sets and a comprehensive parameter estimation methodology. The systematic parameter estimation and uncertainty analysis provides uncertainty information for the single point predictions.

- The developed LFL and UFL model has a higher accuracy than existing GC models and is much simpler to apply than current ANN or QSPR models.
- A temperature-dependent LFL model based on a GC model for a lumped parameter has been developed.
- The advanced parameter estimation using (robust regression) and the systematic outlier treatment using the empirical CDF together with additional experimental data could improve the existing GC MG model for FP and AIT.
- The report of the 95%-confidence interval of the predicted value for the safety-related properties provided important information on the uncertainty (reliability) of the predicted values. The latter is crucial in a quantitative risk assessment as it provides a safety factor for LFL analysis.
- The simplicity of the model application has been demonstrated for the 3-Hexanol as a motivating example.
- The availability of a class of GC models for predicting flammability related properties of chemicals is expected to facilitate the quantitative risk assessment as part of process safety analysis.

References

[1] Gani R, O'Connell JP. Properties and CAPE: From present uses to future challenges. Comput Chem Eng 2001;25:3–14.

[2] Hukkerikar AS, Meier RJ, Sin G, Gani R. A method to estimate the enthalpy of formation of organic compounds with chemical accuracy. Fluid Phase Equilib 2013;348:23–32.

[3] Frutiger J, Marcarie C, Abildskov J, Sin G. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models – an application to heat of combustion. J Chem Eng Data 2016;61:602–13.

[4] Paté-Cornell ME. Uncertainties in risk analysis: Six levels of treatment. Reliab Eng Syst Saf 1996;54:95–111.

[5] Zabetakis MG. Flammability characteristics of combustible gases and vapors. Washington DC: 1965.

[6] Crowl DA, Louvar JF. Definitions of fires and explosions. Chem. Process Saf., Boston: Prentice Hall International Series in the Physical and Chemical Engineering Sciences; 2011, p. 241–3.

[7] Vidal M, Rogers WJ, Holste JC, Mannan MS. A review of estimation methods for flash points and flammability limits. Process Saf Prog 2004;23:47–55.

[8] Rowley JR, Rowley RL, Wilding W V. Estimation of the lower flammability limit of organic compounds as a function of temperature. J Hazard Mater 2011;186:551–7.

[9] Shimy AA. Calculating flammability characteristics of hydrocarbons and alcohols. Fire Technol 1970;6:135–9.

[10] Solovev NV, Baratov AN. Lower limit of flammability of hydrocarbon–air mixtures as a function of the molecular structure of the combustible component. Russ J Phys Chem 1960;34:1661–70.

[11] Oehley E. Ableitung empirischer Gleichungen fur die untere Explosionsgrenze und den Flammpunkt. Chemie Ingenieu Tech 1953;25:399–403.

[12] Shebeko YN, Ivanov A V, Dmitrieva TM. Methods of Calculating the Lower Concentration Limits of Ignition of Gases and Vapors in Air. Sov Chem Ind 1983;15:311–7.

[13] Kondo S, Urano Y, Tokuhashi K, Takahashi A, Tanaka K. Prediction of flammability of gases by using F-number analysis. J Hazard Mater 2001;82:113–28.

[14] Kondo S, Takahashi A, Tokuhashi K. Experimental exploration of discrepancies in F-number correlation of flammability limits. J Hazard Mater 2003;100:27–36.

[15] Seaton WH. Group contribution method for predicting the lower and the upper flammable limits of vapors in air. J Hazard Mater 1991;27:169–85.

[16] R. L. Rowley, W. V. Wilding, J. L. Oscarson, T. A. Knotts, N. F. Giles, DIPPR® Data Compilation of Pure Chemical Properties, Design Institute for Physical Properties, AIChE, New York, NY 2014.

[17] Albahri T a. Flammability characteristics of pure hydrocarbons. Chem Eng Sci 2003;58:3629–41.

[18] Gharagheizi F. Quantitative Structure - Property Relationship for Prediction of the Lower Flammability Limit of Pure Compounds. Energy & Fuels 2008;22:3037–9.

[19] Pan Y, Jiang J, Wang R, Cao H, Cui Y. A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine. J Hazard Mater 2009;168:962–9.

[20] Pan Y, Jiang J, Wang R, Cao H, Cui Y. Prediction of the upper flammability limits of organic compounds from molecular structures. Ind Eng Chem Res 2009;48:5064–9.

[21] Gharagheizi F. A new group contribution-based model for estimation of lower flammability limit of pure compounds. J Hazard Mater 2009;170:595–604.

[22] Albahri T a. Prediction of the lower flammability limit percent in air of pure compounds from their molecular structures. Fire Saf J 2013;59:188–201.

[23] Gharagheizi F. Prediction of upper flammability limit percent of pure compounds from their molecular structures. J Hazard Mater 2009;167:507–10.

[24] Lazzús JA. Neural network/particle swarm method to predict flammability limits in air of organic compounds. Thermochim Acta 2011;512:150–6.

[25] Bagheri M, Rajabi M, Mirbagheri M, Amin M. BPSO-MLR and ANFIS based modeling of lower flammability limit. J Loss Prev Process Ind 2012;25:373–82. doi:10.1016/j.jlp.2011.10.005.

[26] High MS, Danner RP. Prediction of upper flammability limit by a group contribution method. Ind Eng Chem Res 1987;26:1395–9.
[27] Shu G, Long B, Tian H, Wei H, Liang X. Evaluating upper flammability limit of low hydrocarbon diluted with an inert gas using threshold temperature. Chem Eng Sci 2015;138:810–3. doi:10.1016/j.ces.2015.09.013.
[28] Shu G, Long B, Tian H, Wei H, Liang X. Flame temperature theory-based model for evaluation of the flammable zones of hydrocarbon-air-CO2 in mixtures. J Hazard Mater 2015;294:137–44.
[29] Mendiburu AZ, de Carvalho JA, Coronado CR. Estimation of lower flammability limits of C-H compounds in air at atmospheric pressure, evaluation of temperature dependence and diluent effect. J Hazard Mater 2015;285:409–18.
[30] Mendiburu AZ, de Carvalho JA, Coronado CR. Estimation of upper flammability limits of C-H compounds in air at atmospheric pressure, evaluation of temperature dependence and diluent effect. J Hazard Mater 2016;304:512–21.
[31] Zabetakis MG, Lambiris S, Scott GS. Flame temperatures of limit mixtures. 7th Symp. Combust., 1959, p. 484.
[32] Britton LG. Using heats of oxidation to evaluate flammability hazards. Process Saf Prog 2002;21:31–54.
[33] Britton LG, Frurip DJ. Further uses of the heat of oxidation in chemical hazard assessment. Process Saf Prog 2003;22:1–19.
[34] Rowley JR, Rowley RL, Wilding WV. Experimental determination and re-examination of the effect of initial temperature on the lower flammability limit of pure liquids. J Chem Eng Data 2010;55:3063–7.
[35] Catoire L, Naudet V. Estimation of temperature-dependent lower flammability limit of pure organic compounds in air at atmospheric pressure. Process Saf Prog 2005;24:130–7.
[36] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.
[37] Marrero J, Gani R. Group-contribution based estimation of pure component properties. Fluid Phase Equilib 2001;183–184:183–208.
[38] Frutiger J, Abildskov J, Sin G. Outlier treatment for improving parameter estimation of group contribution based models for upper flammability limit. In: Gernaey K V., Huusom JK, Gani R, editors. 12th Int. Symp. Process Syst. Eng. 25th Eur. Symp. Comput. Aided Process Eng., Copenhagen: 2015.
[39] Coward HF, Jones GW. Limits of flammability of gases and vapors. Washington DC: 1952.
[40] Joback K, Reid R. Estimation of pure-component properties from group-contribution. Chem Eng Commun 1987;57:233–43.
[41] Seber G, Wild C. Nonlinear Regression. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1989.
[42] Burnham KP, Andersen DR. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol Methods Res 2004;33:261–304.
[43] Albahri T a. Structural Group Contribution Method for Predicting the Octane Number of Pure Hydrocarbon Liquids. Ind Eng Chem Res 2003;42:657–62.
[44] Kroenlein KC, Chirico R, Diky V, Bazyleva A, Magee J. Thermophysical Property Reliability Issues in the Context of Automated Consumption. 19th Symp. Thermophys. Prop., Boulder, CO: 2015.
[45] Center for Chemical Process Safety. Guidelines for Vapor Cloud Explosion, Pressure Vessel Burst, BLEVE and Flash Fire Hazards. 2nd Editio. Wiley; 2011.
[46] Crowl DA, Louvar JF. Chemical process safety. 2nd ed. New Jersey: Prentice Hall International Series in the Physical and Chemical Engineering Sciences; 2013.
[47] Mills N. ChemDraw Ultra 10.0. J Am Chem Soc 2006;128:13649–50.
[48] Gmehling J, Rasmussen P. Flash Points of Flammable Liquid Mixtures using UNIFAC. Ind Eng Chem Fundam 1982:186–8.

Paper appendix

A.1. Supporting material

The supporting information is permanently available on the website of Journal of Hazardous Materials: http://dx.doi.org/10.1016/j.jhazmat.2016.06.018

Group contribution factors for and formulas for all developed models are shown in tabular form. Furthermore, examples of predicted values including 95% confidence interval for a variety of chemical compounds are given.

The authors developed a software tool where the developed models are implemented. Please contact the corresponding author, Gürkan Sin, for more information.

Paper D

Uncertainty assessment of equations of state with application to an organic Rankine cycle 1

Jérôme Frutiger[a], Ian Bell[b], John P. O'Connell[c], Kenneth Kroenlein[b], Jens Abildskov[a], Gürkan Sin[a]

a) Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
b) Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, USA
c) Department of Chemical Engineering, University of Virginia, Charlottesville, USA

published in Molecular Physics (ISSN: 0026-8976), 115, 1, pages: 1-20, 2017.
DOI: http://dx.doi.org/10.1080/00268976.2016.1275856

Abstract

Evaluations of equations of state (EoS) should include uncertainty. This study presents a generic method to analyse EoS from a detailed uncertainty analysis of the mathematical form and the data used to obtain EoS parameter values. The method is illustrated by comparison of Soave–Redlich– Kwong (SRK) cubic EoS with perturbed-chain statistical associating fluid theory (PC-SAFT) EoS for an organic Rankine cycle (ORC) for heat recovery to power from the exhaust gas of a marine diesel engine using cyclopentane as working fluid. Uncertainties of the EoS input parameters including their corresponding correlation structure, are quantified from experimental measurements using a bootstrap method. Variance-based sensitivity analysis is used to compare the uncertainties from the departure function and the ideal-gas contribution. A Monte Carlo procedure propagates fluid parameter input uncertainty onto the model outputs. Uncertainties in the departure function (SRK or PC-SAFT EoS) dominate the total uncertainties of the ORC model output. For this application and working fluid, SRK EoS has less predictive uncertainty in the process model output than does PC-SAFT EoS, though it cannot be determined if this is due to differences in the data for parameter estimation or in the mathematical form of the EoS or both.

1. Introduction

Low-temperature Organic Rankine Cycles (ORC) systems are used to produce electrical power from waste heat (*e.g.* in marine diesel engine applications [1]). The basic ORC is a power cycle consisting of a pump, an evaporator, an expansion machine and a condenser [2], with a working fluid continuously circulating through the units [3]. In order to evaluate and test promising fluid candidates for a cycle, an equation of state (EoS) is commonly used. In recent years, there has been significant interest in the selection of working fluids for ORCs and optimizing their application.

Screening techniques and multi-criteria database searches [3] as well as Computer Aided Molecular Design (CAMD) [4] have been extensively applied to find appropriate working fluids for ORCs. The reviews by Bao *et al.* [5] on fluid selection, and by Linke *et al.* [6] on molecular fluid design, reference studies concerning working fluids for ORCs.

Several families of EoS have been used for ORC working fluid design and selection studies. Forms of the Helmholtz EoS (as implemented in the well-established REFPROP library [7], or alternatively in the CoolProp library [8]) have been used. The works of Wang *et al.* [9], Chys *et al.* [10], Andreasen *et al.* [11], Zhai *et al.* [12], Luo *et al.* [13], Rödder *et al.* [14], Hærvig *et al.* [15], and Xu *et al.* [16] are examples of screening fluids using high-accuracy fundamental EoS of this form. While libraries like REFPROP or CoolProp implement the most accurate equations of state available in the literature, only a limited number of fluids have been treated with these EoS, preventing wide-range database searches or molecular design studies. However, the rapid development of novel high-performance working fluids that satisfy rigorous safety requirements with low environmental impact [17][18] demand such capabilities.

Cubic EoS, such as Peng-Robinson (PR) [19], Soave-Redlich-Kwong (SRK) [20], and Predictive Soave-Redlich-Kwong (PSRK) [21], have also been used for calculating the thermodynamic properties of ORC working fluids. PR and SRK are particularly convenient for working fluid design studies, because they only require three fluid-specific input properties to their EoS: the critical temperature, T_c, the critical pressure, P_c, and the acentric factor, □. The PR EoS [19] was implemented into molecular design frameworks for working fluids by

Papadopoulos *et al.* [4], [22]. Also, Drescher *et al.* [23], Brown *et al.* [24], Liu *et al.* [25] and Frutiger *et al.* [26] used the PR EoS to screen a large number of working fluids, while Roskosch *et al.* [27] implemented the PR EoS into their reverse engineering methodology for fluid selection. Finally, Sanchez *et al.* [28] predicted the thermodynamic properties of the working fluids in their cycle application with SRK, while Molina-Thierry *et al.* [29] chose PSRK for their CAMD framework.

In addition, Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) [30] has also been used for working fluid properties due to its relatively small number of adjustable parameters that are conceptually related to molecular characteristics. Thus, Lai *et al.* [3][31] applied PC-SAFT to a working fluid analysis of an ORC and a SAFT-type EoS was also used for fluid modeling in the work of Oyeniyi *et al.* [32]. Most recently PC-SAFT was also implemented in a molecular design framework for ORC working fluids by Lampe *et al.* [33].

Additional EoS models have been reported in the literature for prediction of thermophysical properties of working fluids: *e.g.* BACKONE EoS [34], Martin-Hou EoS [35], and Patel-Teja EoS [36]. However, it seems that extended database screening and molecular design for novel fluids is most often performed with either cubic forms, due to their simple structure, or the PC-SAFT EoS with a more complex form but a limited number of parameters.

For ORC applications, an EoS is commonly selected based on goodness-of-fits to data, range of availability of fluid data, limited complexity of model formulation as related to numerical complexity [37], and/or ease of implementation. For example, Kumar *et al.* [38] compared the results of thermodynamic properties obtained from a variety of EoS for gas turbine applications to those from a complex, but highly accurate multi-parameter Helmholtz energy-explicit EoS [39], to determine which simple EOS would best describe compressor efficiency.

In all of these works, when comparisons have been made among different models, there was little concern shown about variations in the number of parameters and their origin from experiment. Analyses were based on the typical application approaches of users who normally select models based on simplicity of form and calculation versus capability to replicate particular experimental data.

However, an additional criterion for the choice of an EoS, that seems not to have been explored thoroughly, is the influence of the uncertainty of the fluid-specific parameters of the EoS on the ORC model output. It should be expected that lower uncertainties would provide more reliable process designs from models.

Experimental property data (*e.g.* critical properties, saturation pressures, and liquid densities) have been normally used to determine parameters of an EoS. These data have associated uncertainties arising from the measurements [40] and how the model incorporates the values [41]. We believe that these property uncertainties should be taken into account when applying an EoS to processes such as ORC [42].

We distinguish the difference between accuracy and uncertainty in the context of computational models for property prediction or process design. Accuracy is the difference between the output predicted by the model and a particular set of experimental measurements of the property or process output. Uncertainty is the range of statistically possible outcomes of the model (usually assumed to be a normal distribution and reported with 95 % confidence). The sources of uncertainty are: 1) the model parameters representing incomplete knowledge of fixed values (input uncertainty); 2) the mathematical formulation of the model only approximating nature (structural uncertainty); and 3) stochastic components of a process

simulation (stochastic uncertainty) [42]. The current study focuses on uncertainties of the parameters of the EoS and their impacts on the uncertainties of ORC process calculations.

In non-linear regression theory, the uncertainties of parameters are defined by the parameter covariance matrix, which should be generated by the developers of the models after parameter optimization. However, developers often do not provide the covariance matrix for EoS studies.

In the preliminary phase of conceptual process design, such as for a new ORC, experimental temperature and pressure data at the process states are often unavailable. Thus, model accuracy and the complementary uncertainty are the only means available to assess potential errors in process design and simulation.

Feistel *et al.* [43] analyzed the uncertainties of empirical reference EoS. They used generalized least squares for parameter regression and propagated the covariance of the input data uncertainties into the calculated values, and into the fitted value covariance matrix. In this way, estimates of the uncertainties of the derived quantities (*e.g.*, the second and third virial coefficients of water) were provided.

Frutiger *et al.* [26] recently presented a methodology to propagate and quantify the impact of parameter uncertainty on an ORC model output, using the PR EoS for thermodynamic properties. A Monte Carlo method was used to propagate the uncertainty of the fluid-specific EoS parameters to the ORC model output. This provided distributions of the cycle power output resulting from fluid property uncertainties. The uncertainties of fluid properties were assumed to be known *a priori*, based on information reported in databases or from information reported in predictive models of pure component properties (*e.g.*, the study of Hukkerikar *et al.* [41]). Several candidate fluids were compared and ranked according to ORC model output uncertainties. This approach allowed the use of uncertainty as an additional dimension in the fluid selection process [26].

A comprehensive methodology to include assessment of model parameter uncertainty based on experimental data is needed. Toward this end, we investigate the following items:

- Quantification of uncertainty and the correlation structure of input properties and parameters based on experimental data
- Sensitivity analysis of the different contributions to the uncertainty of a given EoS, such as ideal-gas versus departure function contributions
- Comparisons of different types of EoS based on fluid-specific uncertainties propagated to the model output of an ORC
- Uncertainty analysis to complement accuracy in selecting an EoS for a given application

We apply a Monte Carlo method for analysis of the commonly used types of EoS in the field of working fluids: cubic (*i.e.*, SRK) and PC-SAFT. Apart from the work of Feistel *et al.* [43], we do not know of any systematic assessments of EoS in terms of uncertainty propagation. We apply this generalized procedure to an ORC application for power generation using a low-temperature heat source of exhaust gas from a marine diesel engine.

The paper is organized as follows: (i) the overall methodology is outlined; (ii) cubic EoS and PC-SAFT, as well as the ORC model formulation, are briefly presented; (iii) the method to obtain the input uncertainties by quantifying experimental error is shown; (iv) the Monte Carlo procedure used to perform uncertainty analysis and variance-based sensitivity analysis is explained; (v) the results of the uncertainty analysis of cubic and PC-SAFT EoS are compared.

2. Method and Tools

The methodology involves the set of steps given in Table 1.

Table 1. Overview of the methodology.

Step 1	**Formulation of EoS and fluid selection**	
Step 2	**Organic Rankine cycle (ORC) process model formulation and optimization of process parameters**	
Step 3	**Quantification of uncertainties in fluid-specific EoS parameters based on experimental data**	
	Step 3.1	Quantification of uncertainty for critical properties and acentric factor for cubic SRK EoS
	Step 3.2	Quantification of uncertainty for parameters of PC-SAFT EoS
	Step 3.3	Quantification of uncertainty for ideal-gas heat capacity parameters
Step 4	**Monte Carlo procedure for input uncertainty propagation to ORC process model output of cubic SRK and PC-SAFT EoS**	
	Step 4.1	Specification of fluid property and parameter input uncertainties
	Step 4.2	Sampling of property and parameter search spaces
	Step 4.3	Evaluation of ORC model for each property and parameter sample
Step 5	**Variance-based sensitivity analysis and EoS selection**	
	Step 5.1	Calculation of variance-based sensitivity measures
	Step 5.2	Analysis and selection of EoS based on accuracy and uncertainty

2.1. Step 1: Formulation of EoS and fluid selection

Models of process cycles require evaluation of thermodynamic properties (*e.g.* enthalpies, entropies, fugacities). The enthalpy, h, and entropy, s, have an ideal contribution (*i.e.* the ideal-gas enthalpy and entropy) and a nonideal gas contribution (departure function, [47]) for the difference between ideal- and real-fluid behaviors:

$$h(P,T) = h^{\mathrm{ref}} + h(T)^{\mathrm{ideal}} + h(P,T)^{\mathrm{dep}} \tag{1}$$

$$s(P,T) = s^{\mathrm{ref}} + s(P,T)^{\mathrm{ideal}} + s(P,T)^{\mathrm{dep}} \tag{2}$$

where P is the pressure and T the temperature. The reference enthalpy and entropy, h^{ref} and s^{ref}, are those defined at the reference state of $T^{\mathrm{ref}} = 273.15$ K and $P^{\mathrm{ref}} = 1\ \mathrm{atm}$. The enthalpy and entropy of the ideal gas at T, are $h(T)^{\mathrm{ideal}}$ and $s(P,T)^{\mathrm{ideal}}$, while $h(P,T)^{\mathrm{dep}}$ and $s(P,T)^{\mathrm{dep}}$ are the respective departure functions. Fugacities can be directly calculated from EoS departure functions, but also more generally from derivatives of the Helmholtz energy

[37][44]. Here we use departure functions from two different EoS for uncertainty analysis: the cubic Soave-Redlich-Kwong (SRK) EoS and the non-associating Perturbed Chain Statistical Association Fluid Theory (PC-SAFT) EoS. The equations differ in mathematical form. A detailed description of the physical background of both cubic and SAFT-type EoS can be found in the work of Kontogeorgis et al. [45].

The SRK EoS originates from Van der Waals-type EoS, in particular the Redlich-Kwong EoS [46]. The underlying principle of van der Waals EoS is to improve upon the ideal-gas law by including attractive and repulsive terms. Soave [20] extended the Redlich-Kwong EoS, by making the *a* parameter a function of temperature and the acentric factor, □,

$$P = \frac{R\,T}{V_{\mathrm{m}} - b} - \frac{a\,\alpha}{V_{\mathrm{m}} \cdot (V_{\mathrm{m}} - b)} \tag{3}$$

In Eq. (3) R is the universal gas constant, T is the absolute temperature, *P* is the absolute pressure and V_{m} is the molar volume. Soave defined the □□function as:

$$a = \frac{0.42747\,R^2\,T_{\mathrm{c}}^2}{P_{\mathrm{c}}} \tag{4}$$

$$b = \frac{0.08664\,R\,T_{\mathrm{c}}}{P_{\mathrm{c}}} \tag{5}$$

$$\alpha = \left(1 + (0.480 + 1.574 \cdot \omega - 0.176 \cdot \omega^2) \cdot \left(1 - \sqrt{T_{\mathrm{r}}}\right)\right)^2 \tag{6}$$

$$T_{\mathrm{r}} = \frac{T}{T_c} \tag{7}$$

Eq. (7) defines T_{r} as used in Eq. (6). Thus, knowing the three primary properties T_c, P_c and ω for a fluid, its departure thermodynamic properties can be calculated from the SRK EoS. The formulations for fugacity, enthalpy, and entropy can be found in Soave [20] and Poling, *et al.* [47]. The Peng-Robinson EoS [19] is a cubic model closely related to SRK and performs similarly for any given process model. The computational implementation of SRK EoS was done by Liu et. al [25]. The uncertainty propagation of Peng-Robinson EoS has been investigated by Frutiger *et al.* [26]

Statistical Associating Fluid Theory (SAFT), is based on a statistical thermodynamic theory for fluids with a repulsive core and directional short-range attractive sites. Economou [48] has reviewed the development of SAFT-type EoS. The Perturbed chain-SAFT (PC-SAFT) EoS for non-associating fluids [30] treats molecules as chains of spherical elements with a pair potential. A temperature-dependent hard sphere diameter *d(T)* for the segments is used to describe the soft repulsion of molecules [49]

$$d(T) = \sigma[1 - 0.12 \cdot \exp(-3\epsilon/(kT))] \tag{8}$$

In Eq. (8) σ is the segment diameter (size parameter), ϵ is the depth of the intersegment molecular pair potential (energy parameter), and *k* is the Boltzmann constant.

In the PC-SAFT EoS, thermodynamic properties become a sum of a hard chain contribution and perturbation terms according to the second-order perturbation theory of Barker and Henderson [49]. Thus, the nonideal Helmholtz energy, A_{res} of a system of *N* chain molecules has the form

$$\frac{A_{\mathrm{res}}}{NkT} = \frac{A_{\mathrm{hc}}}{NkT} + \frac{A_{\mathrm{disp}}}{NkT} \tag{9}$$

where A_{hc} is the hard-chain reference contribution and A_{disp} is the dispersion contribution. The detailed expressions for all thermodynamic properties can be found in the work of Gross *et al.* [30]. In addition to σ, and ϵ, a chain length parameter, *m*, is included. The energy parameter is generally reported as ϵ/k. The computational implementation of PC-SAFT EoS is based on the work of Gross *et al.* [30] and of Fakouri Baygi *et al.* [50].

Both SRK and PC-SAFT require three fluid-specific parameters. However, SRK uses properties (T_c, P_c and ω), which can be measured. Typically T_c is determined directly, while P_c and ω are obtained from vapor pressure curves [51]. The PC-SAFT parameters (σ, ϵ/k and *m*) must be obtained by fitting the EoS to a combination of property data, *e.g.*, vapor pressure and (liquid) density data as functions of temperature [30]. Experimental data used to determine the EoS parameters are often subject to non-negligible uncertainties [40], so this needs to be included in any uncertainty analysis.

The ideal-gas enthalpy and entropy terms are obtained by integrating a temperature-dependent ideal-gas heat capacity function, $c_p(T)$, with parameters obtained from fitting thermal or spectroscopic measurements combined with molecular theory. We use the Aly-Lee ideal-gas heat capacity form with five compound-specific input parameters (*A*, *B*, *C*, *D*, *E*) [52].

$$c_{\mathrm{p}}(T) = A + B \cdot \left[\frac{\frac{C}{T}}{\sinh\left(\frac{C}{T}\right)} \right]^2 + D \cdot \left[\frac{\frac{E}{T}}{\cosh\left(\frac{E}{T}\right)} \right]^2 \tag{10}$$

For the present fluid, cyclopentane, fluid property data were obtained from NIST ThermoData Engine [53][54] for T_c, P_c, □; the DIPPR 801 AIChE database [55] for *A*, *B*, *C*, *D*, *E;* and from Gross and Sadowski [30] for σ, ϵ/k, and *m*.

Natural refrigerants, such as cyclopentane show promising performance in Organic Rankine cycles, have no ozone depletion potential, and possess much lower global warming potential compared to fluorinated and chlorinated compounds [56] some of which are being phased-out in Europe [17]. The disadvantage of natural refrigerants is that many, including cyclopentane, are highly flammable. The input property and parameter data of cyclopentane are listed in Table 2.

Table 2. EoS input property and parameter data for cyclopentane.

Input properties to cubic SRK	T_c [K]	P_c [Pa]	ω [-]			Ref.
	511.7	4.5110^6	0.19			[67]
Input parameters to PC-SAFT	σ [Å]	ϵ/k [J/K]	m [-]			Ref.
	3.7114	265.83	2.3655			[30]
Input parameters to Aly-Lee heat capacity	A [J/(kmol·K)]	B [J/(kmol·K)]	C [K]	D [J/(kmol·K)]	E [K]	Ref.
	41600	301400	1462	180950	669	[55]

We have compared the accuracy of SRK and PC-SAFT for cyclopentane with a reference EoS [57]. For the calculation of saturation pressure as function of temperature from 290 K to 510 K, the PC-SAFT EoS had an average relative error of 0.05%, while SRK had 0.20%. Hence, although both agree well, the PC-SAFT EoS was found to be more accurate, at least for saturation pressure. However, this is not unexpected, because PC-SAFT parameters were fitted to vapour pressure data. We have not compared the results for liquid densities.

2.2. Step 2: Organic Rankine cycle (ORC) model formulation and optimization of process parameters

The Organic Rankine Cycle (ORC) process of this study is a waste heat recovery (WHR) system for electricity production on a large container ship [58]. The process model is based on the work of Andreasen *et al.* [59] and Frutiger *et al.* [26]. Frutiger *et al.* provided a detailed description of the process model.

In the ORC process, the exhaust gas of an on-board MAN diesel engine provides the high temperature heat, with the low temperature heat rejected to sea water [60]. The ORC system has five main components: pump, evaporator (preheater, evaporator and superheater), turbine, condenser, and recuperator (see Figure 1). The working fluid is saturated liquid at low pressure at the pump inlet (state 1). The pump pressurizes it to state 2. It is then directed through the recuperator, to heat up the stream to state 3. It enters the evaporator for preheating to the saturated liquid state 4, evaporating and (optionally) superheating to state 5. In the turbine, the hot pressurized vapor expands to state 6, producing mechanical power which is converted to electricity by a generator connected to the turbine. The low pressure vapor condenses in the recuperator (state 7) and in the condenser completes the cycle to state 1.

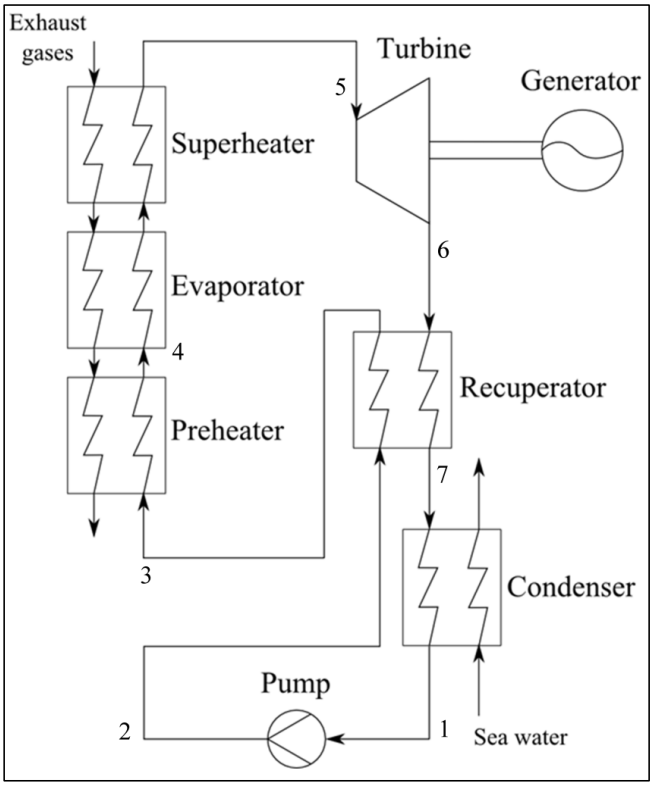


Figure 1. An overview over the ORC process adapted from Andreasen et al. [59].

The process data were provided by MAN Diesel and Turbo [61]. The modelling constraints of the process and of the hot fluid are summarized in Table 3. Engine exhaust gas (*i.e.* air), at a temperature of 222 °C and mass flow rate of 95.4 kg/s, serves as the heat source. Further constraints are: 1) the exhaust gas (air) outlet temperature is limited to 160 °C; 2) at the turbine inlet and outlet, as well as at the saturated liquid point, the minimum temperature difference between the hot and cold streams in the evaporator is set to 10 K; 3) the whole cycle operates at subcritical conditions with the maximum evaporator pressure limited to 0.95 P_c.

Table 3. Constraints for the ORC process model.

Process parameter	*Value*
Exhaust gas (hot fluid) inlet temperature	222 °C
Exhaust gas (hot fluid) outlet temperature	160 °C
Exhaust gas (hot fluid) mass flow rate	95.4 kg/s
Exhaust gas (hot fluid) pressure	0.11 MPa
Condensation temperature	30 °C
Condenser outlet vapor quality (state 1)	0

Pump isentropic efficiency	0.8
Minimum evaporater temperature difference	10 K
Minimum recuperator temperature difference	10 K
Turbine isentropic efficiency	0.8
Minimum turbine outlet vapor quality (state 6)	1

The assumptions used in the numerical modeling are: No pressure losses in piping or heat exchangers, no heat loss from the system, and steady state operation [26].

The outputs from the ORC process model are the net power output W_{NET}, the mass flow $\dot{m}_{\mathrm{wf}}$ of the working fluid, and state variables such as pressures P_i, temperatures T_i, entropies s_i, and enthalpies h_i, (see Figure 1). The net power output W_{NET} (i.e. the difference between turbine power production and pump power consumption) can be calculated from Eq.(11).

$$W_{\mathrm{NET}} = W_{\mathrm{Turb}} - W_{\mathrm{Pump}} \tag{11}$$

$$W_{\mathrm{Turb}} = \dot{m}_{\mathrm{wf}}(h_5 - h_6) \tag{12}$$

$$W_{\mathrm{Pump}} = \dot{m}_{\mathrm{wf}}(h_2 - h_1) \tag{13}$$

where h_i is the enthalpy at state *i* (see Figure 1) and $\dot{m}_{\mathrm{wf}}$ is the mass flow rate of the working fluid, given by energy balances over the evaporator, preheater, and superheater:

$$\dot{m}_{\mathrm{wf}} = \dot{m}_{\mathrm{air}} \cdot \frac{c_{\mathrm{p\ air}}(T_{\mathrm{air\ in}} - T_{\mathrm{air\ out}})}{h_5 - h_3} \tag{14}$$

In Eq. (14) $c_{\mathrm{p\ air}}$ is the heat capacity of the hot air (exhaust gas), which is assumed constant; $T_{\mathrm{air\ in}}$, the temperature of the air flowing into the ORC (*i.e.* the output temperature of the diesel engine); $T_{\mathrm{air\ out}}$, the temperature of air leaving the ORC; and $\dot{m}_{\mathrm{air}}$, the exhaust gas mass flow rate.

The thermal efficiency of the cycle can be expressed as

$$\eta_{\mathrm{therm}} = \frac{W_{\mathrm{NET}}}{\dot{m}_{\mathrm{air}} c_{\mathrm{p\ air}}(T_{\mathrm{air\ in}} - T_{\mathrm{air\ out}})} \tag{15}$$

A degrees of freedom analysis of the cycle suggests that two process variables can be solved for and optimized. We choose the turbine inlet pressure, P_5, and the turbine inlet temperature, T_5. The optimal process conditions were identified by performing particle swarm optimization [62] for cyclopentane.

2.3. Step 3: Quantification of uncertainty of fluid-specific EoS parameters based on experimental data

The goal of this step is to obtain the uncertainties and the correlation matrix of the cubic SRK input parameters (T_{c}, P_{c}, ω), the PC-SAFT parameters (σ, ϵ/k, m), and the Aly-Lee heat capacity parameters (A, B, C, D, E). The quantification is based on the thermodynamic property data. In order to achieve this, the bootstrap method described by Efron [63] is used.

The bootstrap method attempts to quantify the underlying distributions of residual errors commonly defined in statistical contexts as the differences between the experimental data and their corresponding model calculations. This should not be confused with the thermodynamic term "residual function", which is related to the thermodynamic departure function [47]. The residual errors are used to obtain synthetic data sets for parameter estimation by using random sampling with replacement. This procedure is a form of nonlinear propagation of measurement errors to errors as parameter estimators. It is different from non-linear regression theory which relies on asymptotic approximation of the parameter covariance matrix that requires calculation of the jacobian matrix and the assumption that measurement errors are independently identically distributed and follow normal distribution with means equal to zero [64]. In many practical application, this assumption is rarely met (see for instance the residual plots in Hukkerikar *et al.* [41]. Therefore, the bootstrap method that works with the actual distribution of residuals is more appropriate to use in such situations. The method has previously been applied to the development of group contribution methods by Frutiger *et al.* [65].

We now outline the bootstrap method [63]. A generic model $F([\theta])$ with parameters $[\theta]$ to predict variable y^{pred} is given by

$$y^{\text{pred}} = F([\theta]) \tag{16}$$

The goal is to fit the model parameters giving y^{pred} to the experimental data set, y^{exp}, of N_{data} data points, obtaining the parameter estimates θ^* and their corresponding uncertainties.

(1) A reference parameter estimation is made using a non-linear least squares method to obtain the first parameter estimates $[\theta^*]$:

$$\theta^* = \arg\min \sum_i (y_i^{\text{exp}} - F_i(\theta))^2 \tag{17}$$

(2) The residual error for each data point is defined as:

$$\hat{\epsilon}_i = (y_i^{\text{exp}} - y_i^{\text{pred}}) \tag{18}$$

Each residual error $\hat{\epsilon}_i$ has equal probability of occurring, with a probability of $1/N_{\text{data}}$.

(3) New synthetic data sets are produced via the bootstrap method. Random sample replacements are made of residual errors $\hat{\epsilon}$ to generate *k* synthetic data sets (*y**(1); *y**(2), …, *y**(*k*)), each with N_{data} data points. In practice, this bootstrap method simply samples errors and adds them randomly to the estimated properties in the reference step above (*i.e.*, it rearranges the errors):

$$y_i(k)^* = y_i^{\text{pred}} + \hat{\epsilon}_i \qquad \hat{\epsilon}_i \in \hat{F}(\epsilon) \tag{19}$$

where *i* (from 1 to N_{data}) stands for the index of measured data and $\hat{F}$ is the probability function of $\hat{\epsilon}$ (with probability of realization of $1/N_{\text{data}}$ for all $\hat{\epsilon}$)

(4) The least squares parameter estimation is repeated using each synthetic data set *y**(*k*), which results in a new set of estimated parameters □*(*k*) and a new set of predicted values, $y^{\text{pred}*}(k)$. In this way, distributions of the parameters as well as of the predicted values are obtained for representing the uncertainty in the estimated values.

(5) Inference statistics can be used to estimate the mean and standard deviation (SD) of the distributions:

$$SD(\theta^*) = \sqrt{\frac{1}{N_{\text{data}} - 1} \sum_{k=1}^{n} (\theta^*(k) - \mu_{\theta^*})^2} \tag{20}$$

In Eq. (20), $\theta^*(k)$ are the estimated parameters from the k^{th} synthetic data set and μ_{θ^*} is its mean value, which is given by

$$\mu_{\theta^*} = \frac{1}{N_{\text{data}}} \sum_{k=1}^{n} \theta^*(k) \tag{21}$$

The obtained standard deviations are estimates of the parameter uncertainties.

Another important feature of the bootstrap method is that it allows estimation of the correlation structure between the errors of the different parameters (*e.g.* for PC-SAFT: the correlation structure between the residual errors associated with values of σ, ϵ/k and *m*). It is essential for the uncertainty analysis (Step 4 of the methodology) to preserve the original correlation structure, to avoid the output variance calculation being incorrect [66].

Quantification of uncertainty for critical properties and acentric factor for cubic SRK EoS

For many hydrocarbons, T_c has been measured experimentally [47][67]. Hence, its measurement uncertainty serves as input uncertainty for the EoS in this study. However, P_c and ω are often obtained from vapor pressure curves as described by Patel and Ambrose [51][47]. As an example, the Antoine equation [68] can be used:

$$P_c = P_{\text{sat}}(T_c) = 10^{[A_{\text{Ant}} + B_{\text{Ant}}/(C_{\text{Ant}} + 0.1T_c)]} \tag{22}$$

$$\omega = -\log_{10}\left(\frac{10^{[A_{\text{Ant}} + B_{\text{Ant}}/(C_{\text{Ant}} + 0.7T_c)]}}{P_c}\right) - 1 \tag{23}$$

where A_{Ant}, B_{Ant}, and C_{Ant} are the respective Antoine parameters.

Experimental data for the vapor pressure as a function of temperature for the working fluid cyclopentane were taken from the literature [69]. Afterwards a bootstrap method, as described above, was applied: 1) the experimental vapor pressure curve was fitted to an initial set of parameters; 2) new synthetic data sets were generated by random sampling of the errors; and 3) parameter estimation was repeated using each synthetic data set and subsequently P_c and ω were calculated at T_c. Then T_c was perturbed within its stated measurement uncertainty [67]. In order to propagate the measurement errors in temperature to other experimentally measured variables, the Monte Carlo procedure was used. In the Monte Carlo method, 150 random samples from the measurement errors of the temperature were taken and for each sample, variables (A_{Ant}, B_{Ant}, C_{Ant}) were calculated using Eq. (22) and Eq. (23). For the Antoine model, Eq. (23) is usually reliable for □, but Eq. (22) may not be very good for P_c. However, for estimating uncertainties both should be adequate.

The uncertainties of P_c and ω are defined as two standard deviations ($2SD$) of the distributions obtained by the bootstrap method. This is an engineering standard to account for uncertainty with 95% confidence. Figure 2 shows the distribution of T_c, P_c, and ω as obtained from the

bootstrap method; the forms are similar to normal distributions. A summary of results obtained by the bootstrap method for all of the parameters is given in Table 7. The correlation structure was obtained by calculating the correlation matrix of the errors of T_c, P_c, and ω (see Table 4). The cubic EoS parameters were highly correlated (*i.e.* the elements of the correlation matrix were larger than 0.7). The estimated uncertainties in T_c, P_c, and □ are given in Table 7.

Table 4. Correlation matrix of errors of T_c, P_c, and ω from the bootstrap method.

	T_c	P_c	ω
T_c	1		
P_c	0.96	1	
ω	-0.93	-0.85	1

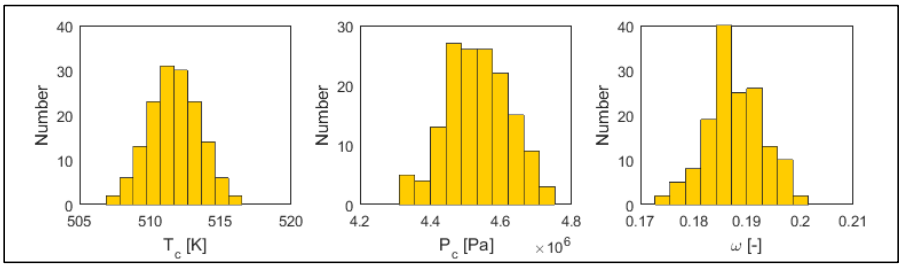


Figure 2. Distribution of SRK parameters from the bootstrap method.

Quantification of uncertainty for parameters of PC-SAFT EoS

The PC-SAFT parameters are usually obtained by fitting residual functions of PC-SAFT [30] to vapor pressure and saturated liquid density data. However, Gross and Sadowski [30] did not report uncertainties of σ, ϵ/k and *m*, for use in our uncertainty propagation analysis.

As a result, we applied the bootstrap method using collected experimental data for vapor pressure [69] over the temperature range of 230-350 K and saturated liquid densities [70] for a temperature range of 190-310 K. Following the methodology as outlined above, 1) the experimental data were fitted to the PC-SAFT EoS; 2) new synthetic data sets were obtained; and 3) parameter estimation was repeated (with 150 random samples) using each synthetic data set. The uncertainties of σ, ϵ, and *m* were obtained by calculating the standard deviation of the respective distributions, and the correlation structure was calculated through the matrix of errors of σ, ϵ/k, and *m* (see Table 5). The parameters ϵ/k and *m* were highly correlated, but σ was not strongly correlated with the other parameters. Figure 3 shows the distribution of σ, ϵ/k, and *m* as obtained from the bootstrap method. The distributions are only roughly in normalized form. The estimated uncertainties for σ, ϵ/k, and *m* can be found in Table 7.

Table 5. Correlation matrix of errors of σ, ϵ/k, and *m* from the bootstrap method.

	σ	ϵ/k	m
σ	1		
ϵ/k	0.05	1	
m	-0.36	-0.94	1

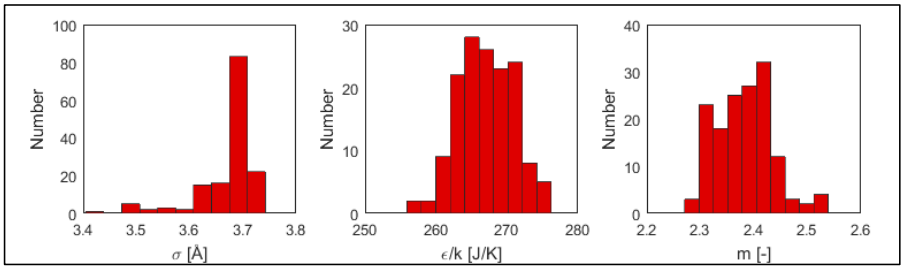


Figure 3. Distribution of PC-SAFT parameters from bootstrap method.

Quantification of uncertainty for ideal-gas heat capacity parameters

The bootstrap method was also applied to obtain the uncertainties and the correlation structure of the respective Aly-Lee heat capacity parameters from experimental data [71]. As for the examples above, the standard deviation of the respective bootstrap-derived distributions for parameters *A*, *B*, *C*, *D*, *E* quantified the uncertainties, and the matrix of errors allowed for the calculation of the correlation structure (see Table 6). With the exception of parameter *B*, all heat capacity parameters were highly correlated with each other. The quantified input uncertainties for *A*, *B*, *C*, *D*, *E* can be found in Table 7. Figure 4 shows the distribution of *A*, *B*, *C*, *D*, *E* as obtained from the bootstrap method; these are similar to a normalized distribution.

Table 6. Correlation matrix of errors of *A*, *B*, *C*, *D*, *E* from the bootstrap method.

	A	*B*	*C*	*D*	*E*
A	1				
B	0.28	1			
C	0.92	0.63	1		
D	0.96	0.51	0.99	1	
E	0.99	0.40	0.96	0.99	1

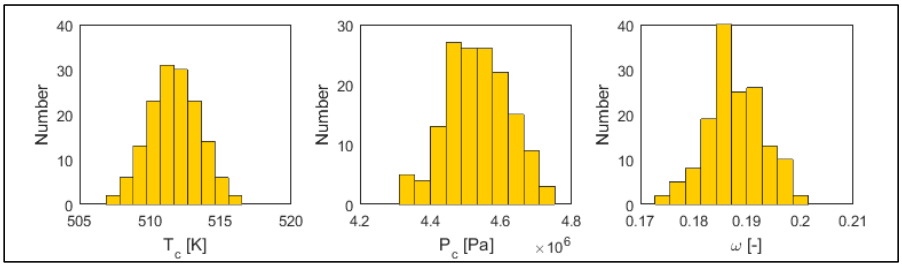


Figure 4. Distribution of heat capacity parameters from the bootstrap method.

Table 7. Estimated uncertainties for the respective SRK, PC-SAFT, and heat capacity parameters in %, as calculated from the ratio between calculated two standard deviations (*SD*) and the actual value from the literature.

Uncertainties in cubic SRK EoS	$\frac{2 \cdot SD(T_c)}{T_c}$		$\frac{2 \cdot SD(P_c)}{P_c}$		$\frac{2 \cdot SD(\omega)}{\omega}$	Ref.
	0.70 %*		3.82 %		5.65 %	[67]
Uncertainties in PC-SAFT EoS	$\frac{2 \cdot SD(\sigma)}{\sigma}$		$\frac{2 \cdot SD(\epsilon/k)}{\epsilon/k}$		$\frac{2 \cdot SD(m)}{m}$	Ref.
	3.05 %		2.89 %		4.61 %	[30]
Uncertainties in Aly-Lee heat capacity model	$\frac{2 \cdot SD(A)}{A}$	$\frac{2 \cdot SD(B)}{B}$	$\frac{2 \cdot SD(C)}{C}$	$\frac{2 \cdot SD(D)}{D}$	$\frac{2 \cdot SD(E)}{E}$	Ref.
	0.34 %	0.46 %	0.79 %	0.61 %	0.34 %	[55]

*directly from experimental measurement uncertainty

2.4. Step 4: Monte Carlo procedure for input uncertainty propagation to ORC model output of cubic SRK and PC-SAFT EoS

A Monte Carlo procedure was used to propagate uncertainties in the fluid-specific EoS parameters to the ORC model output. The procedure follows the work of Frutiger *et al.* [26] as summarized below.

Specification of fluid property and parameter input uncertainties

The quantified uncertainties of the fluid parameters (from Step 3) serve as input uncertainties to be propagated through the ORC model. We do not intend to improve the accuracy of primary property or parameter data values. On the contrary, we use the reported parameter values of Table 2 together with the estimated uncertainties.

Sampling of property and parameter search spaces

Sampling is the key step of the Monte Carlo procedure. The Latin Hypercube Sampling method [72] was utilized for probabilistic sampling of 250 values from the fluid property parameter input space of each EoS. The respective uncertainty defined the range of each

property parameter. The parameters were assumed to be distributed as found in Step 3. The calculated correlations between the respective parameters were taken into account using the rank-based method for correlation control of Iman and Conover [66]. For the obtained heat capacity constants (*A*, *B*, *C*, *D*, *E*), the sampling procedure was performed twice, once with the SRK parameters (T_c, P_c, ω) and a second time with the PC-SAFT parameters (σ, ϵ/k, *m*).

Evaluation of ORC model for each property and parameter sample

The ORC model was evaluated for each of the 250 input property parameter samples resulting from Step 4.2. The ORC model simulations for the SRK EoS were carried out as follows:

(1) The sample sets for the heat capacity input A, B, C, D and E (input for the ideal-gas contribution) and the SRK input properties (T_c, P_c, □) were evaluated together.
(2) The heat capacity parameters were kept constant, while every sample for the SRK input properties was evaluated.
(3) The SRK input properties were kept constant and every sample for the heat capacity input parameters was evaluated.

This procedure was repeated with the samples from the PC-SAFT parameters and the sample set of heat capacity parameters. In this way, it was possible to quantify the influence of the model output uncertainties caused by the ideal-gas and residual functions contributions. Furthermore, the uncertainty propagations of SRK could be directly analyzed, and compared with the ones of PC-SAFT.

Although the error quantification by the bootstrap method and the Monte Carlo procedure were applied only to the SRK and PC-SAFT EoS in this study, the approach is completely general, and can be applied to any type of EoS to analyze the propagation of input uncertainty to the output of the EoS model.

The uncertainty analysis was implemented in Matlab (Mathworks, R14) [73]. The software for performing the uncertainty analysis can be provided as m-script files upon request to the corresponding author.

2.5. Step 5: Variance-based sensitivity analysis and EoS selection

The results of the Monte Carlo uncertainty propagations were distributions of the model outputs (*e.g.* the net power output of the ORC W_{NET}). The broader a model output distribution is, the more uncertain is the model output value. The variance of a distribution is a measure of its width and can be used to quantify output uncertainties, subject to the property uncertainties. Given the distribution of a variable from the Monte Carlo sample evaluation, the associated variance of the distribution can be defined. For W_{NET} this is

$$VAR(W_{\mathrm{NET}}) = \frac{1}{n-1}\sum_{i=1}^{n}\left(W_{\mathrm{NET}}(i) - \mu_{W_{\mathrm{NET}}}\right)^2 \tag{24}$$

where $W_{\mathrm{NET}}(i)$ is the net power output of one Monte Carlo simulation, *n* is the number of simulations, and $\mu_{W_{\mathrm{NET}}}$ is the mean value of the distribution, defined as

$$\mu_{W_{\mathrm{NET}}} = \frac{1}{n}\sum_{i=1}^{n} W_{\mathrm{NET}}(i) \tag{25}$$

The standard deviation is the square root of the variance,

$$SD(W_{\mathrm{NET}}) = \sqrt{VAR(W_{\mathrm{NET}})} \tag{26}$$

To compare the different uncertainty propagations, subject to the EoS parameter uncertainties, a variance-based sensitivity analysis was performed. Sensitivity analysis yields the impact of model parameter uncertainty on the model output uncertainties [74].

Step 5.1: Calculation of variance-based sensitivity measures

The influence of different uncertainty sources on the ORC model outputs may be analyzed by comparing the different variances and standard deviations. As a result, we can compare ideal-gas contributions to uncertainties with those from the nonideal departure functions, and SRK can be considered relative to PC-SAFT. In order to facilitate such comparisons, a sensitivity measure is useful. An example is the sensitivity measure described by Saltelli *et al.* [75] for the net power output W_{NET} of the ORC.

First, we denote the variance of the specific distribution of W_{NET} that results from only the input uncertainties of the SRK EoS (keeping heat capacity parameters constant) by $VAR\left(W_{\mathrm{NET}}(SRK|\,\overline{C_{\mathrm{p}}})\right)$. Then, the variance of the specific distribution of W_{NET} that results from input uncertainties of both the SRK EoS and the heat capacity parameters is denoted $VAR\left(W_{\mathrm{NET}}(SRK, C_{\mathrm{p}})\right)$. The sensitivity measure for SRK input properties, S_{SRK}, with respect to the model output uncertainties is then given by

$$S_{SRK} = \frac{VAR\left(W_{\mathrm{NET}}(SRK|\,\overline{C_{\mathrm{p}}})\right)}{VAR\left(W_{\mathrm{NET}}(SRK, C_{\mathrm{p}})\right)} \tag{27}$$

Eq. (27) quantifies the influence of a propagated input property uncertainty of the SRK EoS on the overall propagated uncertainty. Similarly, the sensitivity measure for the influence of other input parameter uncertainties (heat capacity, PC-SAFT parameters) to other ORC model output properties (*i.e.*, enthalpies, entropies, temperatures and pressures at different stages) can be evaluated.

Step 5.2: Analysis and selection of EoS based on uncertainty

Based on the sensitivity measures and the distributions of the model outputs from the Monte Carlo simulations, we can address the following questions:

(1) Do input uncertainties originating from the ideal-gas contribution or from the departure functions have stronger influence on the model output?
(2) Which of the two departure function input uncertainties (SRK or PC-SAFT) has the stronger effect on the model output?
(3) Which of the two departure functions (SRK or PC-SAFT) has a lower standard deviation in the ORC model output uncertainty and, consequently, might be preferred from the standpoint of process uncertainty?

3. Results and Discussion

The results are now presented as follows: (1) an overview of the uncertainty analysis results; (2) the sensitivity of the ideal-gas contribution compared to the sensitivity of the departure functions (residual functions); and (3) the uncertainty of SRK compared to that of PC-SAFT.

3.1. Overview of the output uncertainties in log(P)-h and T-s diagrams

The outcome of the Monte Carlo methods is shown on temperature-entropy (*T*-*s*) and logarithmic pressure-enthalpy (log(*P*)-*h*) diagrams in Figure 5. The uncertainty is a varying band for both the saturation curves (yellow) and the cycle design (red). All the simulation results obtained from each single property parameter sample are overlaid. The solid black line represents the mean values of the model outputs. From a statistical point of view, the uncertainty bands correspond to the distribution of the model outputs and directly show the sensitivities with respect to the fluid property values. The larger the width of the band, the greater the uncertainty. Hence, Figure 5 gives an overview of all the uncertainty analyses for the SRK EoS (left hand side) and the PC-SAFT EoS (right hand side). These results are analyzed in more detail in the following sections. Figure 5 also gives an overview of the different ORC model outputs that have been further considered for sensitivity analysis in the following results sections (see also Table 8).

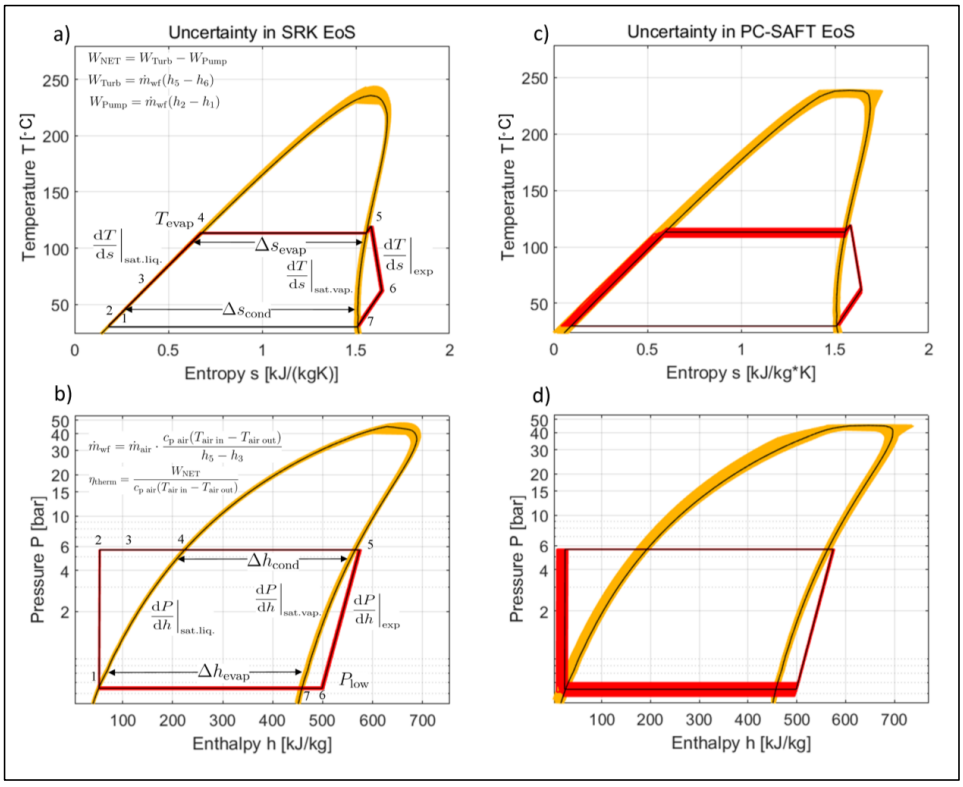


Figure 5. Representation of uncertainty with respect to the fluid properties in the *T*-*s* diagram and log(*P*)-*h* diagram for cyclopentane for SRK and PC-SAFT input uncertainty: Monte Carlo simulations overlaid (yellow/red) and mean (solid black line). The numbers refer to the states of the ORC cycle according to Figure 1. Table 8 lists the symbols for model outputs.

Table 8. Considered model outputs.

ORC net power output	W_{NET} [kW]
Turbine output	W_{Turb} [kW]
Pump work input	W_{Pump} [kW]
Thermal efficiency	η_{therm} [-]
Mass flow of the working fluid	$\dot{m}_{\mathrm{wf}}\left[\frac{\mathrm{kg}}{\mathrm{s}}\right]$
Evaporation temperature	T_{evap} [K]
Lower pressure level	P_{low} [kPa]
Condensation entropy	$\Delta s_{\mathrm{cond}}\left[\frac{\mathrm{kJ}}{\mathrm{kgK}}\right]$
Condensation enthalpy	$\Delta h_{\mathrm{cond}}\left[\frac{\mathrm{kJ}}{\mathrm{kg}}\right]$
Evaporation entropy	$\Delta s_{\mathrm{evap}}\left[\frac{\mathrm{kJ}}{\mathrm{kgK}}\right]$
Evaporation enthalpy	$\Delta h_{\mathrm{evap}}\left[\frac{\mathrm{kJ}}{\mathrm{kgK}}\right]$
Slope of the expansion line in log(*P*)-*h* diagram	$\left.\frac{\mathrm{d}P}{\mathrm{d}h}\right\|_{\mathrm{exp}}\left[\frac{\mathrm{kPa}\cdot\mathrm{kg}}{\mathrm{kJ}}\right]$
Slope of the expansion line in *T*-*s* diagram	$\left.\frac{\mathrm{d}T}{\mathrm{d}s}\right\|_{\mathrm{exp}}\left[\frac{\mathrm{K}^2\cdot\mathrm{kg}}{\mathrm{kJ}}\right]$
Slope of the saturated vapor line in log(*P*)-*h* diagram	$\left.\frac{\mathrm{d}P}{\mathrm{d}h}\right\|_{\mathrm{sat.vap.}}\left[\frac{\mathrm{kPa}\cdot\mathrm{kg}}{\mathrm{kJ}}\right]$
Slope of the saturated vapor line in *T*-*s* diagram	$\left.\frac{\mathrm{d}T}{\mathrm{d}s}\right\|_{\mathrm{sat.vap.}}\left[\frac{\mathrm{K}^2\cdot\mathrm{kg}}{\mathrm{kJ}}\right]$
Slope of the saturated liquid line in log(*P*)-*h* diagram	$\left.\frac{\mathrm{d}P}{\mathrm{d}h}\right\|_{\mathrm{sat.liq.}}\left[\frac{\mathrm{kPa}\cdot\mathrm{kg}}{\mathrm{kJ}}\right]$
Slope of the saturated liquid line *T*-*s* diagram	$\left.\frac{\mathrm{d}T}{\mathrm{d}s}\right\|_{\mathrm{sat.liq.}}\left[\frac{\mathrm{K}^2\cdot\mathrm{kg}}{\mathrm{kJ}}\right]$

As shown in Figure 5, the critical regions of the PC-SAFT log(*P*)-*h* and *T*-*s* diagrams have large uncertainties due to high sensitivity to the EoS parameter uncertainties. However, since our ORC model was operated subcritically, well away from the critical region, poor modelling

of the critical region should not affect the ORC model outputs. Although the mean of the saturation line (solid line) was modelled smoothly, parameter uncertainty (orange) caused some outliers of this property.

From the overview figures, it is possible to visually analyze the results of the fluid-specific EoS parameter uncertainty propagation. For example, from the output uncertainty from the SRK EoS shown on the *T*-*s* diagram (top of Figure 5), the expansion process uncertainty (states 5 to 6) is larger than the uncertainty in the evaporation line (states 4 to 5). This is also shown in the expansion lines and lower pressure line of the log(*P*)-*h* diagram (bottom of Figure 5). For the PC-SAFT EoS, a comparatively wide band can be seen for the evaporation temperature (states 2 to 5) as well as for the saturated liquid line (states 3 to 4) on the *T*-*s* diagram. Furthermore, the pump (states 1 to 2) and the low pressure process have high uncertainty on the log(*P*)-*h* diagram. Note that the uncertainties of PC-SAFT and SRK cannot be compared directly using Figure 5 because the outputs are normalized by the different EoS mass flow rates. A more appropriate comparison of SRK and PC-SAFT EoS is made below.

3.2. Ideal-gas contribution versus departure function: Comparison of uncertainty propagation of input uncertainties for cyclopentane

The effects of the parameter uncertainties on the ideal-gas contribution (*i.e.*, the heat capacity expression) can be compared to those from the departure functions (*i.e.*, SRK and PC-SAFT). Figure 6 shows the output distributions of the ORC net power output W_{NET} as obtained from the evaluated Monte Carlo samples. The results of the combined uncertainty propagations of the departure functions (SRK and PC-SAFT) and the ideal-gas contributions are shown together with the results from the uncertainty analysis when only the departure functions or the ideal-gas contributions were varied subject to their uncertainties. Figure 6 is divided in two parts: On the left hand side, the propagated input uncertainties of PC-SAFT (red) are compared to the ideal-gas contribution, while on the right hand side, results are shown for SRK (yellow) and the ideal gas contribution. The distributions from PC-SAFT and SRK overlap, though the percentage variations can be considered acceptable. However, the mean value of W_{NET} for the PC-SAFT uncertainty was 2.83 % higher than for the SRK.

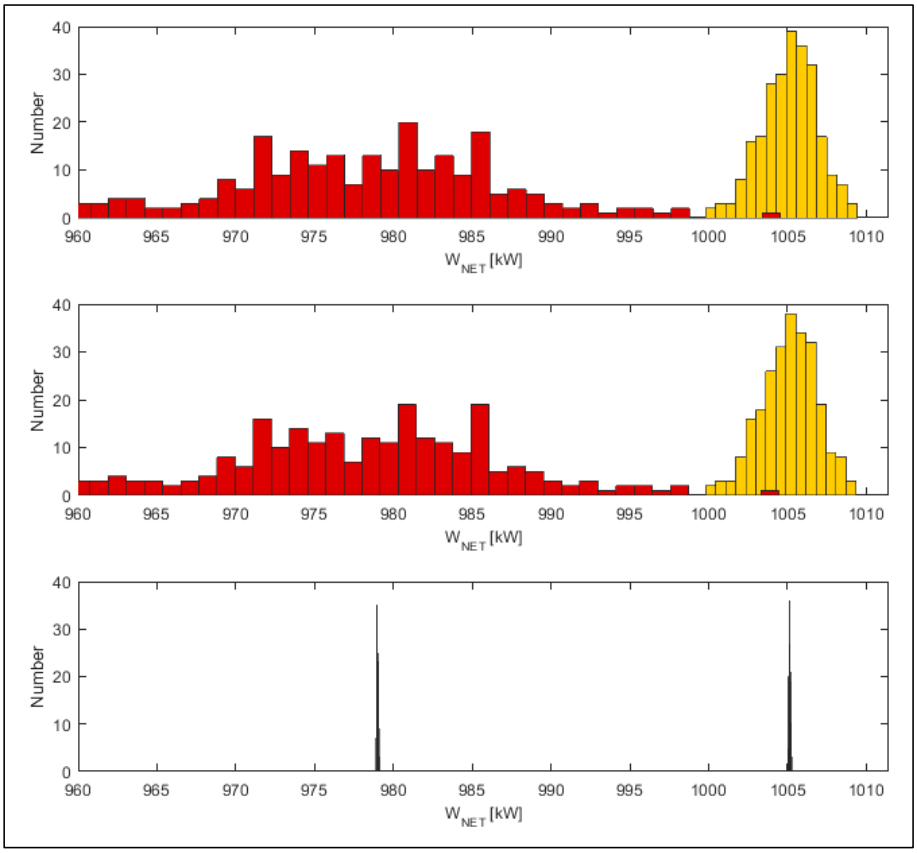


Figure 6. Output distributions of the ORC net power output W_{NET} from Monte Carlo simulations. Subfigures a, b and c compare the output distributions of the propagated input uncertainties of the departure functions SRK (yellow) and PC-SAFT (red) with the ideal-gas contribution (*i.e.*, from heat capacity parameter uncertainties).

Considering the differences in the widths of the distributions of the net power output W_{NET} in Figure 6, the influence of the propagated heat capacity uncertainties on the model output was small compared to the effect of the uncertainties in the departure functions for both SRK and PC-SAFT. The mean of the thermal efficiency of the ORC η_{therm} was 15.05 % for SRK and 14.63 % for PC-SAFT. Hence, it should be noted that the ORC model outputs obtained whith PC-SAFT and SRK do not differ strongly.

Leekumjorn *et al.* [76] thoroughly analyzed the relative errors of both PC-SAFT and SRK compared to experimental values of vapor pressures as functions of temperatures. These authors showed deviations of 2-6 % for a variety of hydrocarbon fluids.

The uncertainty analysis results for other ORC model outputs were analyzed by their respective sensitivity measures, taking into account that the ORC model and the EoS were highly non-linear and the different fluid properties and parameters could potentially influence every model output. Figures 7 (SRK) and 8 (PC-SAFT) give an overview of the results of the uncertainty analysis of all the output variables considered. The sensitivity measures of the input uncertainties from the heat capacity correlation are plotted together with those from the SRK and PC-SAFT EoS.

As Figures 7 and 8 show, the two sensitivities for ideal and nonideal gas contributions sum to unity, because these are additive in the enthalpy and entropy calculations [75].

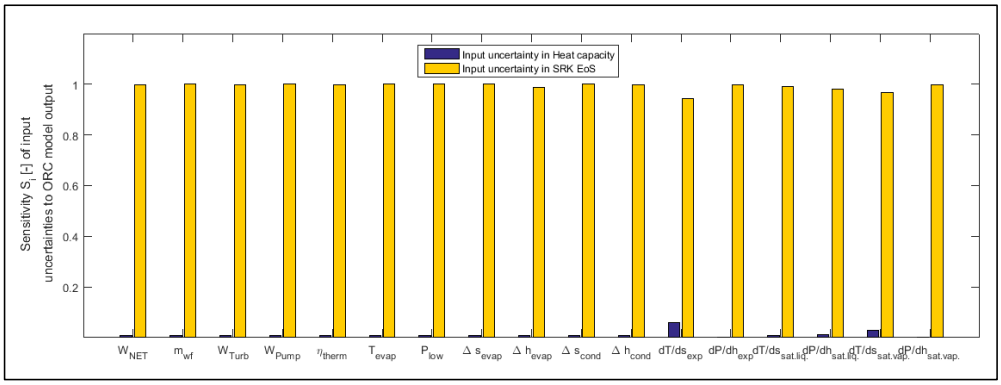


Figure 7. Sensitivity measures S_i for influence of propagated heat capacity parameter uncertainties as well as SRK EoS input uncertainties on the respective model outputs (see also Table 8).

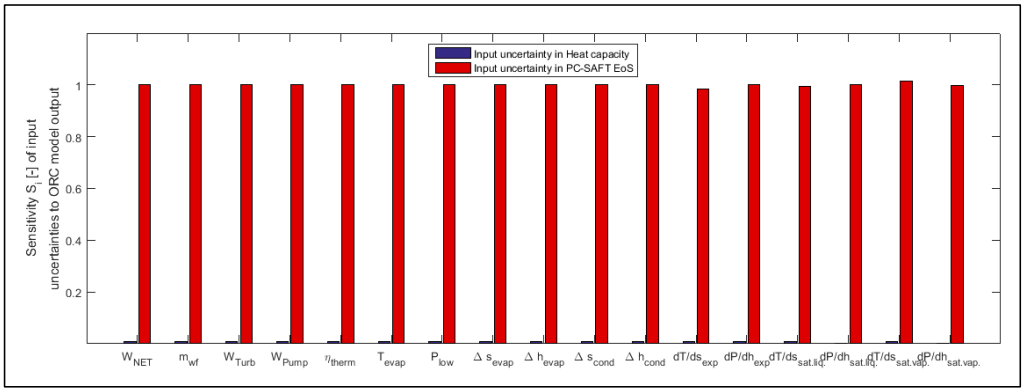


Figure 8. Sensitivity measures S_i for influence of propagated heat capacity parameter uncertainties as well as PC-SAFT EoS input uncertainties on the respective model outputs (see also Table 8).

Sensitivity of the departure functions of SRK and PC-SAFT was much larger than that of the ideal-gas contribution for all the output variables. This is expected since in the ORC both gas and liquid states exist at high pressures. Therefore, the real-gas deviation from the ideal-gas becomes important. Small changes in the ideal-gas enthalpy or entropy contribution do not affect the system strongly, whereas changes of the departure functions will.

There are studies in the literature suggesting that the heat capacity correlation can strongly affect cycle performance [77]. Here, the Aly-Lee heat capacity correlation fitted the experimental data very well over the given temperature range, leading to small uncertainties in the heat capacity parameters (as estimated by the bootstrap method). In addition, the correlation structure was retained. This prevented overestimation of the corresponding uncertainty. The uncertainty in the heat capacity itself was very low (< 1 % uncertainty), which propagates to a small uncertainty in the ideal-gas contribution.

3.3. SRK versus PC-SAFT: Comparison of input uncertainties propagation of and selection of EoS for cyclopentane

In step 1 of Methods and Tools, we compared the accuracy of the two EoS, looking at the differences of experimental and predicted data. The PC-SAFT EoS had an average relative error of 0.05%, while the SRK EoS had 0.20%. Hence, the accuracy of PC-SAFT was superior.

As an additional tool, the SRK and PC-SAFT EoS can be compared in terms of input uncertainty propagation to the ORC model outputs by analyzing the standard deviations of the model output distributions (*e.g.* the distribution of W_{NET} in Figure 6). The standard deviations of the ORC model output distributions for the different ORC model outputs are shown in Figures 9 and 10. Unlike the section before, the sensitivity measure could not be used for the comparison, because the two EoS did not have the same reference variance. Instead the standard deviations of the respective output distributions have been compared.

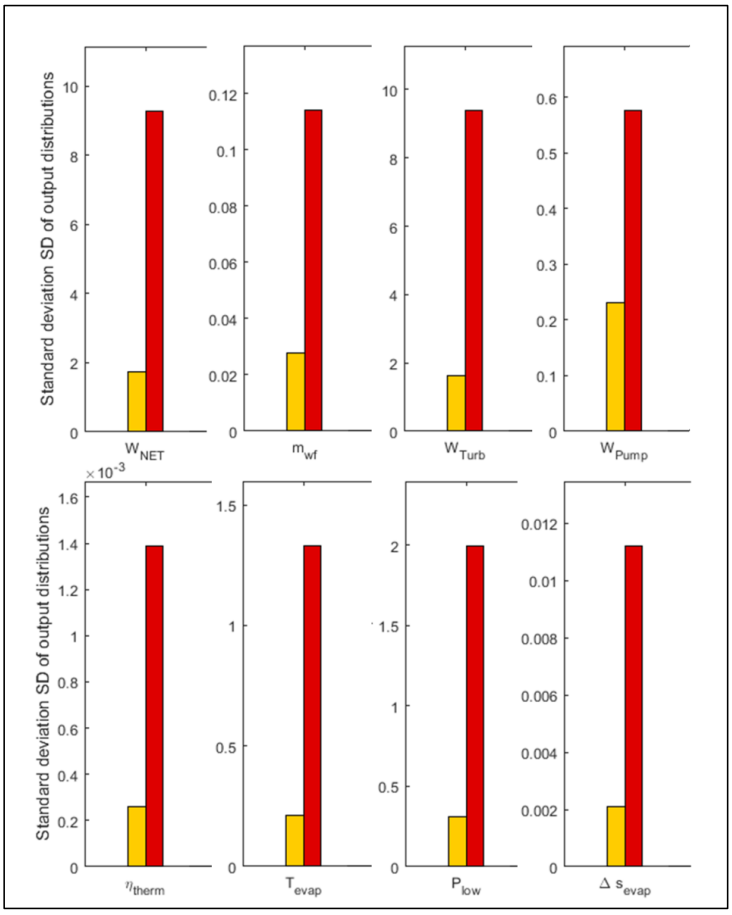


Figure 9. Standard deviations *SD* of the different ORC model output distributions obtained from propagating input parameter uncertainties for SRK (yellow) and input parameters of PC-SAFT (red) (see also Table 8).

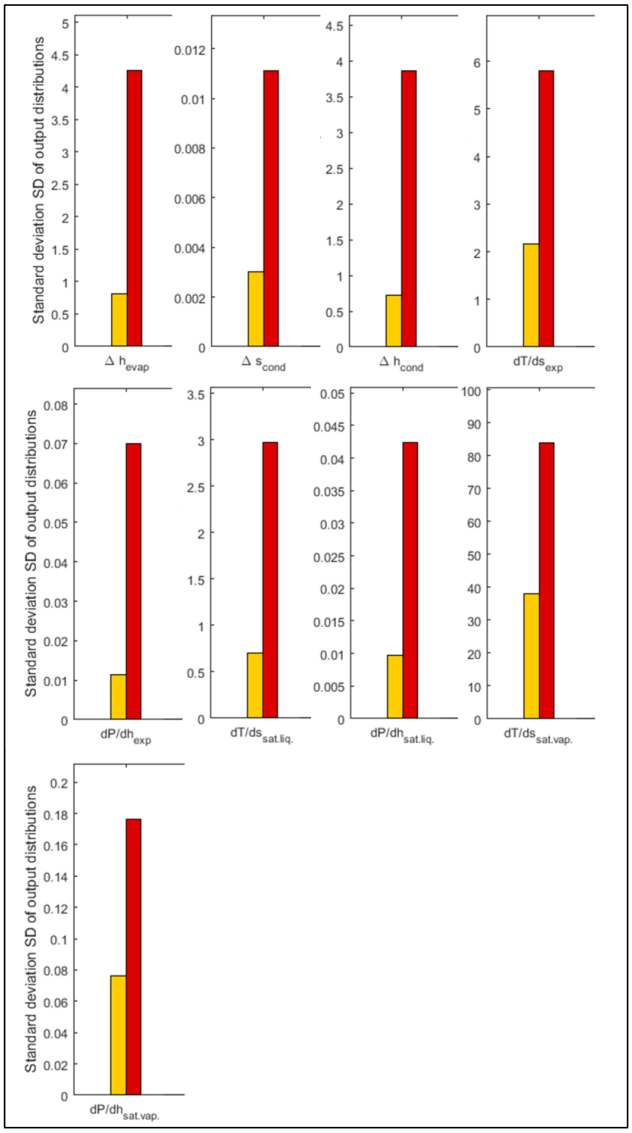


Figure 10. Standard deviations *SD* of the different ORC model output distributions obtained from propagating input parameter uncertainties for SRK (yellow) and input parameters of PC-SAFT (red) (see also Table 8).

The standard deviations of the model output distributions are larger for PC-SAFT. This can also be seen from the width of the distributions of the net power output in Figure 6, which are much larger for PC-SAFT than SRK. However, the parameter uncertainties of SRK and PC-SAFT were similar (see Table 7). Even small uncertainties in the PC-SAFT parameters apparently lead to relatively large output standard deviations, at least compared to the SRK. The uncertainty analysis shows that the uncertainties in the PC-SAFT parameters interact more strongly than do those of the SRK, leading to a higher output uncertainty. The cause of this could be the differences in mathematical form, or the different data used to obtain the parameters, or both. The PC-SAFT parameters enter into several different functions, which are

(from a model point of view) highly nested and often of contrasting effects whereas the SRK parameterization is more direct. In addition, many temperature-dependent data were used to obtain the PC-SAFT parameters while only constant critical property data were used for SRK. Given that the effects are lumped together, it is not possible to separate them.

The Monte Carlo uncertainty analysis used here can be used as an additional criterion to justify the choice of an EoS (in addition to accuracy and computational efficiency). Lower output uncertainties would be desirable from a modeling point of view because the results are expected to be more reliable, especially over extended ranges of conditions. Considering the used experimental data of the thermo-physical properties (i.e. vapour pressure), the SRK EoS is slightly less accurate than PC-SAFT in predicting properties. However, for the present ORC model and working fluid (cyclopentane), one needs to have experimental evaluation of the proposed ORC process design and measure the power output (W_{NET}) in order to calculate the accuracy of the two candidate models: namely ORC model including SRK versus ORC model including PC-SAFT. Such experimental data for ideally more than one working fluid candidates will enable statistical evaluation of accuracy of both models for ORC process design. In the absence of such experimental data, one has the model output uncertainty for both models to work with. As demonstrated in Frutiger et al. [26], the uncertainty in the predicted power output can be used cautiously or optimistically when searching for alternative candidates.

5. Conclusions

Uncertainties of EoS can be analyzed as an additional, complementary tool to EoS accuracy also in situations where experimental data are not available to calculate accuracy. This study developed parameter uncertainties for two types of equations of state (SRK and PC-SAFT) from measured data using a bootstrap method. These EoS parameter uncertainties were propagated via a Monte Carlo procedure to the output of an organic Rankine cycle model for power production via waste heat recovery from the exhaust gas of a marine diesel engine. Variance-based sensitivity analysis allowed for the comparison of the different outcomes of the uncertainty analyses.

It was found that:

- The bootstrap method allowed for the quantification of the uncertainties of the fluid-specific parameters of both EoS, including their corresponding correlation structure, from experimental data.
- The propagated output uncertainties of the ORC model were determined more by uncertainties in the EoS departure functions than uncertainties from the ideal-gas contribution from the heat capacity model.
- The PC-SAFT EoS had an average relative error between experimental and predicted vapor pressure data of 0.05%, while SRK had an error of 0.20%. This suggests that the PC-SAFT EoS seems more accurate. However, this is not unexpected, since PC-SAFT was fitted to a wide range of vapor pressure data on cyclopentane, while SRK was not.
- The range of the ORC model output uncertainties (i.e. the standard deviations of the respective distributions) were smaller for SRK than for PC-SAFT, indicating that, from an uncertainty point of view, the SRK EoS could be preferable for this application, i.e. performance evaluation of working fluid in ORC process design. It cannot be determined if the higher uncertainty of PC-SAFT is due to differences in data for parameter estimation or in the mathematical forms of the EoS. One needs to have experimental evaluation of the proposed ORC process design and measure the power output (W_{NET}) in order to calculate the accuracy of the two candidate models. At this

stage, given that the distribution of uncertainties of PC-SAFT is much broader than that from the SRK, while the property accuracy is not dramatically different, SRK seems preferable.

We suggest that future process modelling studies should examine uncertainty as well as accuracy of potential EoS models in order to gain additional insights about uncertainties in fluid properties, parameters, and EoS model structure. In particular, measurement errors in data should be taken into account when developing and reporting EoS models and the resulting covariance matrix of model parameters should be calculated and reported. This allows direct propagation of parameter uncertainties to model output uncertainties, which provides another and important criterion for property model selection for process design.

References

[1] Bühler F, Fridolin M, Huang B, Andreasen JG, Elmegaard B. Mapping of low temperature heat sources in Denmark. ECOS 2015 28th Int Conf Effic Cost, Optim Simul Environ Impact Energy Syst 2015.

[2] Invernizzi CM. The Organic Rankine Cycle. Closed Power Cycles, Thermodyn. Fundam. Appl., London: Springer Verlang; 2013, p. 117–76.

[3] Lai NA, Wendland M, Fischer J. Working fluids for high-temperature organic Rankine cycles. Energy 2011;36:199–211. doi:10.1016/j.energy.2010.10.051.

[4] Papadopoulos AI, Stijepovic M, Linke P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. Appl Therm Eng 2010;30:760–9.

[5] Bao J, Zhao L. A review of working fluid and expander selections for organic Rankine cycle. Renew Sustain Energy Rev 2013;24:325–42.

[6] Linke P, Papadopoulos A, Seferlis P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. Energies 2015;8:4755–801.

[7] Lemmon E, Huber M, McLinden M. Reference fluid thermodynamic and transport properties-REFPROP, standard reference database 23, version 8.0, National Institute of Standard and Technology; 2007.

[8] Bell I, Wronski J, Quoilin S, Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. Ind Eng Chem Res 2014;53:2498–508.

[9] Wang EH, Zhang HG, Fan BY, Ouyang MG, Zhao Y, Mu QH. Study of working fluid selection of organic Rankine cycle (ORC) for engine waste heat recovery. Energy 2011;36:3406–18. doi:10.1016/j.energy.2011.03.041.

[10] Chys M, Broek M Van Den, Vanslambrouck B, Paepe M De. Potential of zeotropic mixtures as working fl uids in organic Rankine cycles. Energy 2012;44:623–32.

[11] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic rankine cycles. Energy 2014;73:204–13.

[12] Zhai H, Shi L, An Q. Influence of working fluid properties on system performance and screen evaluation indicators for geothermal ORC (organic Rankine cycle) system. Energy 2014;74:2–11.

[13] Luo D, Mahmoud A, Cogswell F. Evaluation of Low-GWP fluids for power generation with Organic Rankine Cycle. Energy 2015;85:481–8. doi:10.1016/j.energy.2015.03.109.

[14] Rödder M, Neef M, Laux C, Priebe KP. Systematic fluid selection for organic Rankine cycles (ORC) and performance analysis for a combined high and temperature low temperature cycle. Proc ASME Turbo Expo 2015 Turbine Tech Conf Expo 2015;138:1–9.

[15] Hærvig J, Sørensen K, Condra TJ. Guidelines for optimal selection of working fluid for an organic Rankine cycle in relation to waste heat recovery. Energy 2016;96:592–602.

[16] Xu J, Yu C. Critical temperature criterion for selection of working fluids for subcritical pressure

Organic Rankine cycles. Energy 2014;74:719–33.
[17] Regulation (EC) No 1005/2009 of the European Parliament and of the council. Off J Eur Communities 2000;L 269:1–15.
[18] McLinden MO, Kazakov AF, Steven Brown J, Domanski PA. A thermodynamic analysis of refrigerants: Possibilities and tradeoffs for Low-GWP refrigerants. Int J Refrig 2014;38:80–92. doi:10.1016/j.ijrefrig.2013.09.032.
[19] Peng D-Y, Robinson DB. A New Two-Constant Equation of State. Ind Eng Chem Fundam 1976;15:59–64.
[20] Soave G. Equilibrium constants from a modified Redlich-Kwong equation of state. Chem Eng Sci 1972;27:1197–203. doi:10.1016/0009-2509(72)80096-4.
[21] Holderbaum T, Gmehling J. PSRK: A Group Contribution Equation of State Based on UNIFAC. Fluid Phase Equilib 1991;70:251–65. doi:10.1016/0378-3812(91)85038-V.
[22] Papadopoulos AI, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward Optimum Working Fluid Mixtures for Organic Rankine Cycles using Molecular Design and Sensitivity Analysis. Ind Eng Chem Res 2013;52:12116–33.
[23] Drescher U, Brüggemann D. Fluid selection for the Organic Rankine Cycle (ORC) in biomass power and heat plants. Appl Therm Eng 2007;27:223–8.
[24] Brown JS, Brignoli R, Daubman S. Methodology for estimating thermodynamic parameters and performance of working fluids for organic Rankine cycles. Energy 2014;73:818–28.
[25] Liu W, Meinel D, Wieland C, Spliethoff H. Investigation of hydrofluoroolefins as potential working fluids in organic Rankine cycle for geothermal power generation. Energy 2014;67:106–16.
[26] Frutiger J, Andreasen JG, Liu W, Spliethoff H, Haglind F, Abildskov J, et al. Working fluid selection for organic Rankine cycles - impact of uncertainty of fluid properties. Energy 2016;109:987–97.
[27] Roskosch D, Atakan B. Reverse engineering of fluid selection for thermodynamic cycles with cubic equations of state, using a compression heat pump as example. Energy 2015;81:202–12. doi:10.1016/j.energy.2014.12.025.
[28] Sánchez D, Monje Brenes B, Muñoz de Escalona, José M. Chacartegui R. Non-conventional combined cycle for intermediate temperature systems. Int J Energy Res 2013;37:403–11.
[29] Molina-Thierry DP, Flores-Tlacuahuac A. Simultaneous optimal design of organic mixtures and rankine cycles for low-temperature energy recovery. Ind Eng Chem Res 2015;54:3367–83.
[30] Gross J, Sadowski G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. Ind Eng Chem Res 2001;40:1244–60. doi:10.1021/ie0003887.
[31] Lai NA, Wendland M, Fischer J. Description of linear siloxanes with PC-SAFT equation. Fluid Phase Equilib 2009;283:22–30. doi:10.1016/j.fluid.2009.05.011.
[32] Oyewunmi OA, Taleb AI, Haslam AJ, Markides CN. An assessment of working-fluid mixtures using SAFT-VR Mie for use in organic Rankine cycle systems for waste-heat recovery. Comput Therm Sci 2014;6:301–16.
[33] Lampe M, Stavrou M, Bücker HM, Gross J, Bardow A. Simultaneous Optimization of Working Fluid and Process for Organic Rankine Cycles Using PC-SAFT. Ind Eng Chem Res 2014;53:8821–30. doi:10.1021/ie5006542.
[34] Saleh B, Koglbauer G, Wendland M, Fischer J. Working fluids for low-temperature organic Rankine cycles. Energy 2007;32:1210–21.
[35] Leck T. Evaluation of HFO-1234yf as a Potential Replacement for R-134a in Refrigeration Applications. Proc 3rd IIR Conf Thermophys Prop … 2009;155:1–9.
[36] Akasaka R, Tanaka K, Higashi Y. Thermodynamic property modeling for 2,3,3,3-tetrafluoropropene (HFO-1234yf). Int J Refrig 2010;33:52–60. doi:10.1016/j.ijrefrig.2009.09.004.
[37] Span R. Multiparameter Equations of State. 2000.
[38] Kumar SK, Kurz R, O'Connell JP. Equations of state for gas compressor design and testing. ASME 1999 Int Gas Turbine Aeroengine Congr Exhib 1999;2:1–10.
[39] Wagner W, Setz U. A new equation of state and tables of thermodynamic properties for methane covering the range from melting line to 625K at pressures up to 1000 Mpa. J Phys Chem Ref Data 1991;20:1061–155.
[40] Dong Q, Chirico RD, Yan X, Hong X, Frenkel M. Uncertainty Reporting for Experimental Thermodynamic Properties. J Chem Eng Data 2005;50:546–50.

[41] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.
[42] Sin G, Gernaey K V., Eliasson Lantz A. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. Biotechnol Prog 2009;25:1043–53.
[43] Feistel R, Lovell-Smith JW, Saunders P, Seitz S. Uncertainty of empirical correlation equations. Metrologia 2016;53:1079–90. doi:10.1088/0026-1394/53/4/1079.
[44] Bell I, Jäger A. Helmholtz energy transformation of common cubic equations of state for use with pure fluids and mixtures. J Res Natl Inst Stand Technol 2016;121:238–63.
[45] Kontogeorgis GM, Folas GK. Industrial Applications Thermodynamic Models for Industrial Applications From Classical and Advanced. West Sussex, UK: John Wiley & Son; 2010.
[46] Redlich O, Kwong JNS. On the thermodynamics of solutions; an equation of state; fugacities of gaseous solutions. Chem Rev 1949;44:233–44. doi:10.1021/cr60137a013.
[47] Poling BE, Prausnitz JM, O'Connell JP. The Estimation of Physical Properties. Prop. Gases Liq., New York: McGraw-Hill; 2004, p. 1–9.
[48] Economou IG. Statistical Associating Fluid Theory: A Successful Model for the Calculation of Thermodynamic and Phase Equilibrium Properties of Complex Fluid Mixtures. Ind Eng Chem Res 2002;41:953–62. doi:10.1021/ie0102201.
[49] Barker JA, Henderson D. Perturbation theory and equation of state for fluids: the square well potential. J Chem Phys 1967;47:2856–61. doi:10.1063/1.1701689.
[50] Fakouri Baygi S, Pahlavanzadeh H. Application of the perturbed chain-SAFT equation of state for modeling CO2 solubility in aqueous monoethanolamine solutions. Chem Eng Res Des 2015;93:789–99. doi:10.1016/j.cherd.2014.07.017.
[51] Ambrose D, Patel NC. The correlation pressures and estimation of vapour IV. Extrapolation of vapour pressures and estimation of critical pressures the principle of corresponding states using two reference fluids with non-spherical molecules. J Chem Thermodyn 1984;16:459–68.
[52] Aly F, Lee L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. Fluid Phase Equilib 1981;6:169–79.
[53] Frenkel M, Chirico RD, Diky V, Yan X, Dong Q, Muzny C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. J Chem Inf Model 2005;45:816–38.
[54] Diky V, Chirico RD, Muzny CD, Kazakov AF, Kroenlein K, Magee JW, et al. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 8. Properties of Material Streams and Solvent Design. Chem Inf Model 2013;53:249–66. doi:10.1021/ci300470t.
[55] R. L. Rowley, W. V. Wilding, J. L. Oscarson, T. A. Knotts, N. F. Giles, DIPPR® Data Compilation of Pure Chemical Properties, Design Institute for Physical Properties, AIChE, New York, NY 2014.
[56] Hwang Y, Ohadi M, Radermacher R. Natural refrigerants. Mech Eng 1998;120:96–9.
[57] Gedanitz H, Davila MJ, Lemmon EW. Speed of sound measurements and a fundamental equation of state for cyclopentane. J Chem Eng Data 2015;60:1331–7. doi:10.1021/je5010164.
[58] Larsen U, Pierobon L, Haglind F, Gabrielii C. Design and optimisation of organic Rankine cycles for waste heat recovery in marine applications using the principles of natural selection. Energy 2013;55:803–12.
[59] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic Rankine cycles. Energy 2014;73:204–13.
[60] MAN Diesel & Turbo. Waste Heat Recovery System (WHRS) for Reduction of Fuel Consumption, Emission and EEDI. 2012. doi:5510-0136-03ppr.
[61] MAN Diesel and Turbo: CEAS Engine Data report 7G95ME-C9.5 with LP SCR & EGB Specified main engine and other parameters. 2016.
[62] Kennedy J. Particle swarm optimization. Encycl Mach Learn 2010;46:685–91. doi:10.1109/ICNN.1995.488968.
[63] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.
[64] Seber G, Wild C. Nonlinear Regression. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1989.
[65] Frutiger J, Marcarie C, Abildskov J, Sin G. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models –

an application to heat of combustion. J Chem Eng Data 2016;61:602–13.
[66] Iman R, Conover W. A distribution-free approach to inducing rank correlation among input variables. Commun Stat Part B-Simulation Comput 1982;11:311–34.
[67] Daubert TE. Vapor - Liquid Critical Properties of Elements and Compounds . 5 . Branched Alkanes and Cycloalkanes. J Chem Eng Data 1996;41:365–72.
[68] Aston JG, Fink HL, Schumann SC. The Heat Capacity and Entropy, Heats of Transition, Fusion and Vaporization and the Vapor Pressures of Cyclopentane. Evidence for a Non-planar Structure. J Am Chem Soc 1943;65:341–6.
[69] Mokbel I, Rauzy E, H. L, C. B, J. J. Vapor pressures of 12 alkylcyclohexanes, cyclopentane, butylcyclopentane and trans-decahydronaphthalene down to 0.5 Pa. Experimental results, correlation and prediction by an equation of state. Fluid Phase Equilib 1995;108:103–20.
[70] Garcia Baonza V, Caceres Alonso M, Nunez Delgado J. Study of the Equation of State of Cyclopentane from 193 to 298 K and Pressures up to 104 MPa. J Phys Chem 1992;96:1859–68.
[71] McCullough JP, Pennington RE, Smith JC, Hossenlopp IA, Waddington G. Thermodynamics of Cyclopentane, Methylcyclopentane and 1 ,cis-3-Dimethylcyclopentane : Verification of the Concept of Pseudorotation. J Am Chem Soc 1959;81:5880–3.
[72] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.
[73] MATLAB 2015b, The MathWorks, Inc., Natick, Massachusetts, United States n.d.
[74] Saltelli A, Sobol IM. About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Saf 1995;50:225–39.
[75] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun 2010;181:259–70. doi:10.1016/j.cpc.2009.09.018.
[76] Leekumjorn S, Krejbjerg K. Phase behavior of reservoir fluids: Comparisons of PC-SAFT and cubic EOS simulations. Fluid Phase Equilib 2013;359:17–23. doi:10.1016/j.fluid.2013.07.007.
[77] Stijepovic MZ, Linke P, Papadopoulos AI, Grujic AS. On the role of working fluid properties in Organic Rankine Cycle performance. Appl Therm Eng 2012;36:406–13. doi:10.1016/j.applthermaleng.2011.10.057.

Paper E

Uncertainty assessment of equations of state with application to an organic Rankine cycle 2

Jérôme Frutiger[a], Ian Bell[b], John P. O'Connell[c], Kenneth Kroenlein[b], Jens Abildskov[a], Gürkan Sin[a]

a) Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
b) Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, USA
c) Department of Chemical Engineering, University of Virginia, Charlottesville, USA

published in Proceedings of the 30th International Conference on Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems *ECOS-2017*, 2017,

and presented at *ECOS-2017*, July 2-6 2017, San Diego, USA.

Abstract

Evaluations of equations of state (EoS) with application to process systems should include uncertainty analysis. A generic method is presented for determining such uncertainties from both the mathematical form and the data for obtaining EoS parameter values. The method is implemented for the Soave–Redlich–Kwong (SRK), the Peng-Robinson (PR) cubic EoS, and the perturbed-chain statistical associating fluid theory (PC-SAFT) EoS, as applied to an organic Rankine cycle (ORC) power system to recover heat from the exhaust gas of a marine diesel engine with cyclopentane as the working fluid. Uncertainties of the EoS input parameters, including their corresponding correlation structure, are quantified from the data using a bootstrap method. A Monte Carlo procedure propagates parameter input uncertainties onto the process output. Regressions have been made of the three cubic EoS parameters from both critical point matching and vapor pressure and density data, as used for the three PC SAFT parameters. ORC power uncertainties of 2-5 % are found for all models from the larger data sets. Mean power values for the cubic EoS are similar for both parameter regressions. The mean power from the PC-SAFT EoS is less than for the cubic EoS, with no overlap of the uncertainty distributions.

1. Introduction

1.1. Uncertainty in equations of state (EoS) and process models

Thermodynamic cycles such as Organic Rankine Cycles (ORCs) and heat pump systems allow for the recovery of waste heat in process industries and converting it into electrical power or supplying heat back into the system. For example, low-temperature waste heat from marine diesel engines can be used to produce electricity to increase the efficiency of the engine and thereby lower fuel costs and CO_2 emissions [1]. In the preliminary and conceptual design phase of new ORCs or heat pumps, process modeling is necessary to plan, analyze and estimate costs for a given application. The proper selection of working fluid is crucial to the performance of the cycle. Screening techniques and Computer Aided Molecular Design (CAMD) have been extensively applied to find appropriate working fluids for thermodynamic cycles [2][3]. In order to evaluate the thermophysical properties (*e.g.*, enthalpy and entropy) of suitable fluids, an appropriate Equation of State (EoS) is used during process simulations [4]. Different families of EoS have been used extensively in modeling thermodynamic cycles, such as forms of the Helmholtz EoS (as implemented in the well- established REFPROP library [5], or alternatively in the CoolProp library [6]), Cubic EoS (such as Peng-Robinson (PR) [7] or Soave-Redlich-Kwong (SRK) [8]) as well as the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) [9]. Typical criteria for the selection of an EoS are goodness-of-fits to data, range of availability of fluid data, limited complexity of model formulation as related to numerical complexity [10], and/or ease of implementation [11]. The study of Kumar *et al*. [11] demonstrated the sensitivity of natural gas compressor efficiencies to EoS modeling. A much less recognized criterion for the choice of an EoS is the influence of the uncertainty of the fluid-specific parameters of the EoS on the respective process model output. The parameters of an EoS are often determined through fitting to experimental property data (*e.g.*, critical properties, saturation pressures, and liquid densities). These data have associated uncertainties arising from the measurements [12] and how the model incorporates the values [13]. When applying an EoS to a process, these property uncertainties propagate into output uncertainty of the corresponding process model [14]. It is important to distinguish between accuracy [15] and uncertainty [14] in the context of computational models for process design. Accuracy is the difference between the output predicted by the model and experimental

measurements of the property or process output. Uncertainty is the range of statistically possible outcomes of the model (usually assumed to be a normal distribution and reported with 95 % confidence). In the preliminary design phase (*e.g.*, of an ORC system) the accuracy of a process model is often unknown as the process has not been physically constructed, so no experimental measurements of process output are available. In the absence of such experimental data, model output uncertainty is a valuable tool to analyze an EoS.

1.2. Uncertainty quantification for EoS

When fitting the property parameters of an EoS to experimental data using non-linear regression methods, the uncertainties of parameters are defined by the parameter covariance matrix. Feistel *et al.* [16] used generalized least squares for parameter regression and propagated the covariance of the input data uncertainties into the calculated values, obtaining the uncertainties of the derived quantities such as the second and third virial coefficients of water. However, developers only rarely provide the covariance matrix for EoS studies. Recently, Frutiger *et al.* [17] presented a Monte Carlo-based methodology to propagate and quantify the impact of property parameter uncertainty on a process model output of an ORC system. Further, this methodology was used to assess and compare the uncertainty propagation for two different types of EoS: Cubic (SRK) and PC-SAFT [18]. The latter study quantified the parameter uncertainty of SRK and PC-SAFT from measured data using a bootstrap method. These EoS parameter uncertainties were then propagated via a Monte Carlo procedure to the output of an ORC model. Variance-based sensitivity analysis allowed for the comparison of the different outcomes of the uncertainty analyses. In particular, the major results were:

- The ORC output uncertainties were dominated by contributions from the EoS departure functions, rather than contributions from ideal gas heat properties.
- The range of the ORC model output uncertainties were smaller for SRK than for PC-SAFT, for the considered ORC application and working fluid [18].

The EoS and its properties were parameterized as recommended in the literature, and its uncertainty was quantified. This means that the SRK EoS parameters were expressed in terms of the critical temperature, T_c, critical pressure, P_c, and acentric factor, ω, so as to ensure the inflection of the critical isotherm at the critical pressure [19] and to (nearly) reproduce the vapor pressure used to obtain the acentric factor. As a consequence, the uncertainties in T_c, P_c, and ω could be determined from measurements of T_c and vapor-pressure data by fitting to an Antoine-equation. The PC-SAFT parameters σ, ε/k, and m were fitted directly to vapor pressure and liquid density data, as suggested by the developers of the EoS [9]. Thus the different approaches to quantifying the uncertainties followed commonly accepted practice in the literature and typical contemporary approaches to EoS application. However, it could not be determined if the different ranges of the output uncertainties were due to differences in the data used for parameter estimation or to the mathematical form(s) or a combination of these. The present study is based on the previous work of Frutiger *et al.* [18]: The property uncertainties are quantified from experimental data and propagated through an ORC model with cyclopentane as a working fluid, providing the ORC net power output for the given application. In particular, we investigate the following items:

- SRK is parameterized by fitting it directly to the same experimental cyclopentane data as PC-SAFT was. This is not the conventional treatment, and a consequence is that the isotherm exhibiting an inflection point will no longer be at the experimental critical temperature/pressure.

- A bootstrap method quantifies the uncertainties of the three parameters (A, B and β - see below), which are then propagated through an ORC model to obtain the ORC output uncertainty.
- The same analysis is made for Peng-Robinson EoS, to analyze the uncertainty propagation of another commonly used cubic EoS and to compare it to SRK and PC-SAFT.
- The sources of uncertainties are investigated in detail to explore whether the data, or the mathematical structure, or both, are of more influence on the output uncertainty.

The paper is structured as follows: (i) the overall methodology as followed by Frutiger *et al*. [18] is outlined; (ii) PC-SAFT, cubic EoS (SRK and Peng-Robinson), as well as the ORC model, are briefly presented; (iii) the results of the uncertainty analysis of cubic and PC-SAFT EoS are compared.

2. Method and Tools

The methodology consists of the major steps given in Table 1 and is based on the work of Frutiger *et al*. [18].

Table 1. Overview of the methodology.

Step 1	Quantification of uncertainties in fluid-specific EoS parameters based on experimental data using Bootstrap method
Step 2	Monte Carlo procedure for input uncertainty propagation to ORC process model output of cubic (SRK and Peng-Robinson) and PC-SAFT EoS
Step 3	Analysis of ORC model output uncertainty distributions

2.1. Quantification of uncertainties in fluid-specific EoS parameters based on experimental data using Bootstrap method

Formulation of EoS

Solving a thermodynamic cycle model requires evaluating conceptual thermodynamic properties (*e.g.*, enthalpies, entropies, fugacities). Enthalpy and entropy have ideal gas contributions and a non-ideal gas contribution (departure function) [25]. Fugacities are directly calculated from EoS departure functions, or more generally from derivatives of the Helmholtz energy. In this study we compare departure functions from cubic and from SAFT-type EoS: Soave-Redlich-Kwong (SRK) EoS, Peng-Robinson EoS and the non-associating Perturbed Chain Statistical Association Fluid Theory (PC-SAFT) EoS. The ideal gas enthalpy and entropy are obtained for all EoS by integrating the ideal gas heat capacity function as described by Aly and Lee [20]. The uncertainties in the ideal gas contributions were described in the work of Frutiger *et al*. [18]. These are not analyzed here.

PC-SAFT is based on a statistical thermodynamic theory for fluids with a repulsive core and directional short-range attractive sites. A temperature-dependent hard-sphere diameter $d(T)$ for the segments is used to describe the soft repulsion of molecules [21]

$$d(T) = \sigma \cdot [1 - 0.12 \cdot \exp(-3\epsilon / (kT))] \tag{1}$$

with σ as the segment diameter (size parameter), ϵ as the depth of the intersegment molecular pair potential (energy parameter, often reported as ϵ/k), and k as the Boltzmann constant. Further, the non-ideal Helmholtz energy, A_{res}, of a system of N chain molecules has the form

$$\frac{A_{res}}{NkT} = \frac{A_{hc}}{NkT} + \frac{A_{disp}}{NkT} \tag{2}$$

with A_{hc} being the hard-chain reference contribution and A_{disp} being the dispersion contribution. Details of the thermodynamic properties of PC-SAFT can be found in the work of Gross *et al.* [9]. In general, the PC-SAFT EoS is always expressed in terms of the parameters σ (segment diameter), ϵ/k (energy parameter), and m (chain length parameter). In the work of Frutiger *et al.* [18], the uncertainties of the PC-SAFT parameters σ, ϵ/k, and m were obtained through fitting to collected experimental data for vapor pressure [22] over the temperature range of 230-350 K and saturated liquid densities [23] for a temperature range of 190-310 K using a Bootstrap method (see next section). The uncertainties in σ, ϵ/k and m were afterwards propagated through an ORC model system to obtain the uncertainty of the ORC model outputs (*i.e.*, the net power output uncertainty). The 3-parameter cubic EoS can be written in the following general form,

$$P = \frac{RT}{V_m - b} - \frac{a\alpha(T)}{V_m^{\,2} + (c+1)V_m b - cb^2} \tag{3}$$

with T being the absolute temperature, P the absolute pressure, V_m the molar volume and R is the universal gas constant. The parameters a, b and c as well as the temperature-dependent function $\alpha(T)$ are specific for the particular version of the cubic EoS [25]. For SRK [8] a, b, c and $\alpha(T)$ are given by

$$a = \frac{0.42747 R^2 T_c^2}{P_c} \tag{4}$$

$$b = \frac{0.08664 R T_c}{P_c} \tag{5}$$

$$c = 0 \tag{6}$$

$$\alpha(T) = \left(1 + \left(0.480 + 1.574 \cdot \omega - 0.176 \cdot \omega^2\right) \cdot \left(1 - \sqrt{T_r}\right)\right)^2 \tag{7}$$

The expressions for a, b and $\alpha(T)$ are developed to guarantee the inflection of the critical isotherm at the critical pressure while T_r is the reduced temperature. The parameters for Peng-Robinson [24] EoS can be similarly expressed as

$$a = \frac{0.45724 R^2 T_c^2}{P_c} \tag{8}$$

$$b = \frac{0.077796 R T_c}{P_c} \tag{9}$$

$$c = 1 \tag{10}$$

$$\alpha(T) = \left(1 + \left(0.375 + 1.542 \cdot \omega - 0.270 \cdot \omega^2\right) \cdot \left(1 - \sqrt{T_\mathrm{r}}\right)\right)^2 \tag{11}$$

In the work of Frutiger *et al*. [18], the property parameter uncertainties in T_c, P_c and ω have been determined and subsequently propagated through the EoS and the process model to the ORC model outputs. For T_c, measurement uncertainty [25][26] served as input uncertainty to the EoS. P_c and ω were obtained using a Bootstrap method to fit vapor pressure to an Antoine equation as described by Patel and Ambrose [27][25]. In this work, we wish to fit the cubic EoS to the same experimental data as used previously for the PC-SAFT EoS and to quantify its uncertainty using a Bootstrap method. This allows comparison uncertainties in the data fitting. To do this, we re-parameterize the two cubic EoS in terms of their parameters *a* and *b* as well as a third parameter β. For SRK the re-parameterized EoS is given by

$$P = \frac{RT}{V_\mathrm{m} - b} - \frac{a\alpha(T)}{V_\mathrm{m}^{\,2} + V_\mathrm{m} b} \tag{12}$$

$$\alpha(T) = \left(1 + \beta \cdot \left(1 - \sqrt{T_\mathrm{r}}\right)\right)^2 \tag{13}$$

A similar formulation was used for Peng-Robinson EoS. In this way we do not ensure the isotherm with an inflection to be at the critical temperature; instead we consider *a*, *b* and β as fluid-specific parameters that should be obtained by fitting the EoS to experimental data. The result is the same number of regressed parameters for all of the EoS models.

Quantification of EoS parameter uncertainty using bootstrap method

The uncertainties and the correlation matrix of the corresponding EoS property parameters are from thermodynamic property data of Daubert [26]. The detailed description and mathematical formulation of the uncertainty quantification can be found in the work of Frutiger *et al*. [18]. The bootstrap method as described by Efron [28] is used. The method quantifies the underlying distributions of residual errors commonly defined in statistical contexts as the differences between the experimental data and their corresponding model calculations. The residual errors are then used to produce synthetic data sets for use in parameter estimation by using random sampling with replacement. In the current study, uncertainties in the PR and SRK EoS parameters were obtained by fitting the EoS to experimental data and carrying out a bootstrap method to obtain the parameter distribution. Vapor pressures [22] over the temperature range of 230-350 K and saturated liquid densities [23] for a temperature range of 190-310 K of cyclopentane have been used. The key steps of the bootstrap methods are the following:

1. A reference parameter estimation is carried out using a non-linear least-squares method to obtain the first parameter estimates.

2. The residual error (*i.e.*, the difference between the experimental and predicted value) for each data point is calculated.

3. New synthetic data sets are produced by *bootstrapping*: residual errors are sampled and added randomly to the estimated properties in the reference step above (*i.e.*, re-arranging the errors).

4. The least squares parameter estimation is repeated using each synthetic data set.

5. The obtained distribution of parameters is analyzed by interference statistics (mean and standard deviation).

6. Uncertainties of the respective parameters are defined as two standard deviations (2*SD*) of the distributions obtained by the bootstrap method, which is an engineering standard to account for uncertainty with 95 % confidence.

Figure 1 shows the distributions of the EoS parameters of SRK and Peng-Robinson EoS as obtained from the bootstrap method. The breadths of the distributions are similar with the SRK being slightly broader. Because of the different model constructions, the differences in the distribution of the parameters is to be expected.

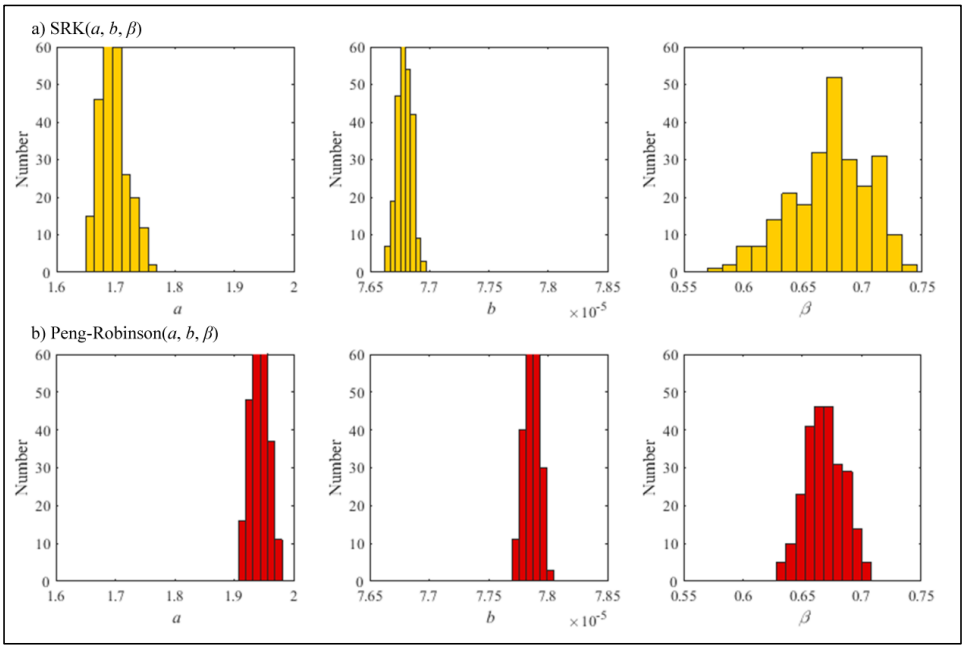


Figure 1. Distribution of SRK and Peng-Robinson parameters *a*, *b* and β obtained from the bootstrap method.

Table 2 summarizes the quantified uncertainty results in *a*, *b* and β.

Table 2. Estimated parameters and uncertainties for SRK and Peng-Robinson parameters, respectively. The uncertainties as calculated from the ratio between calculated two standard deviations (SD) and the actual value from the literature.

		a	*b*	*β*
SRK	Ref. values [26]	1.72	8.17×10^{-5}	0.78
	Mean values of distribution	1.70	7.69×10^{-5}	0.67
	Uncertainties	2.72 %	0.16 %	9.80 %
		a	*b*	*β*
Peng-Robinson	Ref. values [26]	1.84	7.34×10^{-5}	0.66
	Mean values of distribution	1.94	7.79×10^{-5}	0.67
	Uncertainties	1.56 %	0.68 %	4.87 %

2.2. Uncertainty propagation of fluid-specific EoS parameters through ORC model

2.2.1. ORC process formulation

The quantified parameter uncertainties of the corresponding EoS are propagated through an ORC application for power generation using a low-temperature heat source of exhaust gas from a marine diesel engine. Cyclopentane is the working fluid. The process model is based on the work of Andreasen *et al.* [29]. The detailed model description and equations can be found in Frutiger *et al.* [17][18]. Figure 2 gives an overview over the system containing the components and the corresponding modeling constraints of the process and of the hot fluid.

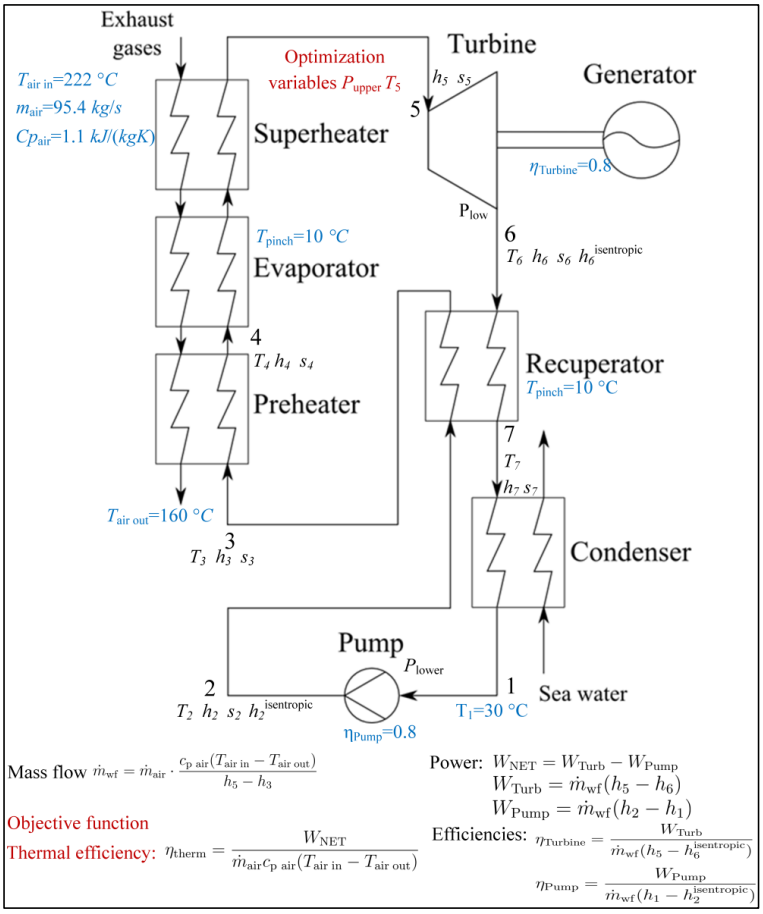


Figure 2. An overview over the ORC process adapted from Frutiger *et al.* [18]. The objective function is the thermal efficiency η_{therm}, which is optimized subject to the objective variables P_{high} and T_5 and the specified process parameters.

The ORC layout has five main components: pump, evaporator (preheater, evaporator and superheater), turbine, condenser, and recuperator. The outputs from the ORC process model are the net power output W_{NET}, the mass flow $\dot{m}_{wf}$ of the working fluid, and state variables such as pressures P_i, temperatures T_i, entropies s_i, and enthalpies h_i, (see Figure 2). According to a degrees-of-freedom analysis of the cycle, two process variables can be solved for and optimized. The turbine inlet pressure, $P_5 = P_{high}$, and the turbine inlet temperature, T_5 have been selected. The optimal process conditions were identified by optimization for cyclopentane.

Monte Carlo procedure for parameter uncertainty propagation through ORC model output of cubic (SRK and Peng-Robinson) and PC-SAFT EoS

A Monte Carlo procedure was used to propagate uncertainties in the fluid-specific EoS parameters to the ORC model power output, W_{NET}. The Monte Carlo method is based on the work of Frutiger *et al.* [17][18] and is as follows:

1. Specification of fluid property and parameter input uncertainties: The quantified uncertainties of the fluid parameters serve as input uncertainties to be propagated through the ORC model.

2. Monte Carlo sampling of property and parameter search spaces: Latin Hypercube Sampling method [30] is used for probabilistic sampling of 250 values from the fluid property parameter defined the range of each property parameter uncertainty. The obtained correlations between the respective parameters were taken into account using the rank-based method for correlation control of Iman and Conover [31].

3. Evaluation of ORC model for each property and parameter sample: The ORC model is evaluated for each of the 250 input parameter samples resulting from Step 2.

In the study of Frutiger *et al*. [18], this procedure was carried out for PC-SAFT parameterized in terms of σ, ϵ/k and m and for SRK EoS parameterized in T_c, P_c and ω. Here, the procedure is applied for SRK and Peng-Robinson EoS parameterized in terms of a, b, and β as well as for the Peng-Robinson EoS parameterized in T_c, P_c and ω. The present results can be compared to the previous results

3. Results and discussion

Figure 3 shows the output distributions of the ORC net power output W_{NET} as obtained from the evaluated Monte Carlo samples. The results of the combined uncertainty propagations of the departure functions of SRK, Peng-Robinson and PC-SAFT EoS are shown in the three parts: The upper subfigure shows the distribution of PC-SAFT parameterized in σ, ϵ/k, and m, along with the cubic EoS (SRK and Peng-Robinson) parameterized in T_c, P_c and ω. The middle and the bottom subfigures depict the distributions of SRK and of Peng-Robinson parameterized in a, b and β. Furthermore, Table 3 shows the mean values of the distributions with their corresponding uncertainties.

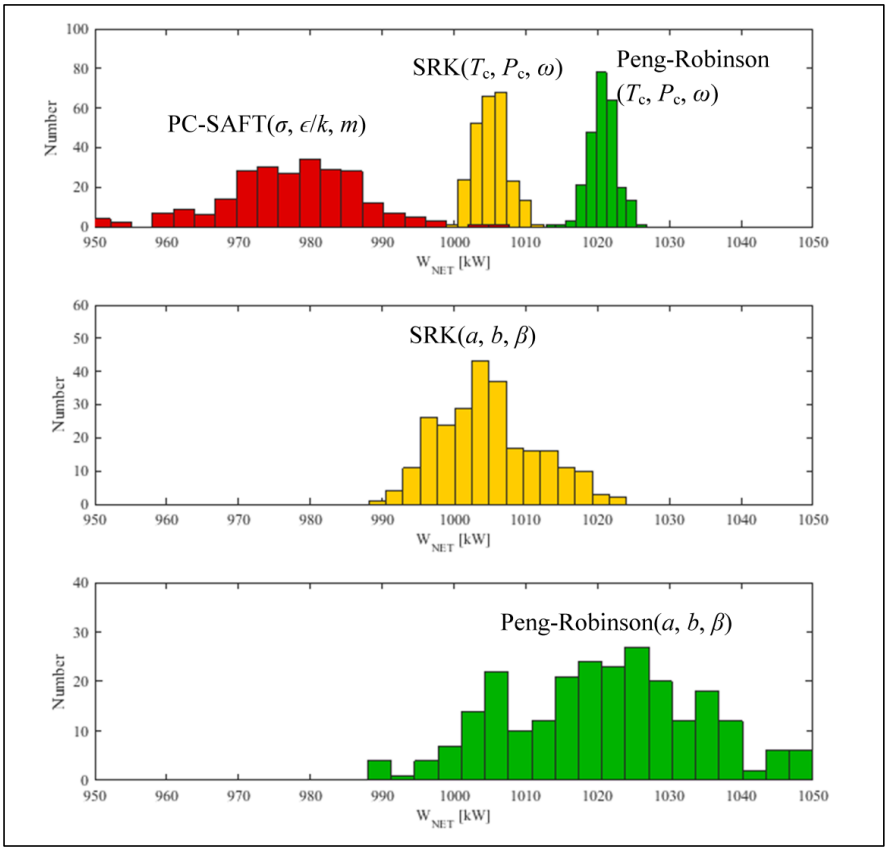


Figure 3. Output distributions of the ORC net power output W_{NET} from Monte Carlo simulations. The subfigures compare the output distributions of the propagated input uncertainties of the departure functions SRK (yellow), PC-SAFT (red) and Peng-Robinson (green). Distributions of SRK and Peng-Robinson EoS are shown when parameterized in both (T_c, P_c and ω) and (a, b and β).

Table 3. Output uncertainties for W_{NET} subject to the propagation of the respective input uncertainties of SRK, Peng-Robinson and PC-SAFT parameters. The uncertainties as calculated from the ratio between calculated two standard deviations (SD) and the mean value (95 %-confidence). Results from the study of Frutiger et al. [18] are marked with *.

	Mean values of distribution of W_{NET}	Uncertainties
PC-SAFT*	976 kW	1.94 %
SRK (T_c, P_c, ω)*	1005 kW	0.47 %
Peng-Robinson(T_c, P_c, ω)	1021 kW	0.38 %
SRK (a, b, β)	1005 kW	1.36 %
Peng-Robinson (a, b, β)	1020 kW	3.18 %

Figure 3 and Table 3 are revealing in several aspects. First, the two parameterizations of the cubic EoS give different uncertainty distributions. The (a, b, β) forms are much broader than those from (T_c, P_c, ω), and, similar in breadth to that for the PC-SAFT, are slightly more sharp about the mean. This difference seems to be due to the greater amount of data used in that regression. For practical purposes, the uncertainties in W_{NET} are significant (of the order of 2 – 5 %), and need to be recognized when designing processes such as an ORC. Second, the EoS models give different mean values for W_{NET} with the PC-SAFT giving the lowest by 5 %. There is significant overlap of the distributions only for the (a, b, β) forms of the cubic EoS. Thus, the process results do depend on the model form. Third, the mean values for the cubic EoS are essentially the same for both parameterizations. The later insights suggest that the propagation from properties to mean process outcomes is determined more by model form than by data used in regressions.

4. Conclusions

The uncertainty propagation of different EoS has been investigated for cubic and PC-SAFT EoS. The EoS parameter uncertainties were quantified from measured data using a bootstrap method. The uncertainties were propagated through an ORC cycle model to obtain the uncertainty range of the net power output subject to the uncertainty of the EoS parameters. The common parameterizations of the SRK and PR EoS models yielded somewhat narrower uncertainty distributions than PC-SAFT, with higher net process power values. When the cubic EoS models were reformulated with three regressed parameters, and fitted to the same data as with the PC-SAFT, the uncertainty distributions for power became much broader, though the mean values where quite similar and still greater than those from PC-SAFT. The effects of uncertainties on the power in the example Organic Rankine Cycle are 2-5 %.

References

[1] Bühler F, Fridolin M, Huang B, Andreasen JG, Elmegaard B. Mapping of low temperature heat sources in Denmark. ECOS 2015 28th Int Conf Effic Cost, Optim Simul Environ Impact Energy Syst 2015.

[2] Bao J, Zhao L. A review of working fluid and expander selections for organic Rankine cycle. Renew Sustain Energy Rev 2013;24:325–42.

[3] Linke P, Papadopoulos A, Seferlis P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. Energies 2015;8:4755–801.

[4] Saleh B, Koglbauer G, Wendland M, Fischer J. Working fluids for low-temperature organic Rankine cycles. Energy 2007;32:1210–21.

[5] Lemmon E, Huber M, McLinden M. Reference fluid thermodynamic and transport properties-REFPROP, standard reference database 23, version 8.0, National Institute of Standard and Technology; 2007.

[6] Bell I, Wronski J, Quoilin S, Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. Ind Eng Chem Res 2014;53:2498–508.

[7] Peng D-Y, Robinson DB. A New Two-Constant Equation of State. Ind Eng Chem Fundam 1976;15:59–64.

[8] Soave G. Equilibrium constants from a modified Redlich-Kwong equation of state. Chem Eng Sci 1972;27:1197–203. doi:10.1016/0009-2509(72)80096-4.

[9] Gross J, Sadowski G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. Ind Eng Chem Res 2001;40:1244–60. doi:10.1021/ie0003887.

[10] Span R. Multiparameter Equations of State. 2000.

[11] Kumar SK, Kurz R, O'Connell JP. Equations of state for gas compressor design and testing. ASME 1999 Int Gas Turbine Aeroengine Congr Exhib 1999;2:1–10.
[12] Dong Q, Chirico RD, Yan X, Hong X, Frenkel M. Uncertainty Reporting for Experimental Thermodynamic Properties. J Chem Eng Data 2005;50:546–50.
[13] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.
[14] Sin G, Gernaey K V., Eliasson Lantz A. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. Biotechnol Prog 2009;25:1043–53.
[15] Joint Committee for Guides in Metrology. International vocabulary of metrology — Basic and general concepts and associated terms (VIM). 2008.
[16] Feistel R, Lovell-Smith JW, Saunders P, Seitz S. Uncertainty of empirical correlation equations. Metrologia 2016;53:1079–90. doi:10.1088/0026-1394/53/4/1079.
[17] Frutiger J, Andreasen JG, Liu W, Spliethoff H, Haglind F, Abildskov J, et al. Working fluid selection for organic Rankine cycles - impact of uncertainty of fluid properties. Energy 2016;109:987–97.
[18] Frutiger J, Bell I, O'Connell JP, Kroenlein K, Abildskov J, Sin G. Uncertainty assessment of equations of state with application to an organic Rankine cycle. Mol Phys 2017.
[19] Redlich O, Kwong JNS. On the thermodynamics of solutions; an equation of state; fugacities of gaseous solutions. Chem Rev 1949;44:233–44. doi:10.1021/cr60137a013.
[20] Aly F, Lee L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. Fluid Phase Equilib 1981;6:169–79.
[21] Barker JA, Henderson D. Perturbation theory and equation of state for fluids: the square well potential. J Chem Phys 1967;47:2856–61. doi:10.1063/1.1701689.
[22] Mokbel I, Rauzy E, H. L, C. B, J. J. Vapor pressures of 12 alkylcyclohexanes, cyclopentane, butylcyclopentane and trans-decahydronaphthalene down to 0.5 Pa. Experimental results, correlation and prediction by an equation of state. Fluid Phase Equilib 1995;108:103–20.
[23] Garcia Baonza V, Caceres Alonso M, Nunez Delgado J. Study of the Equation of State of Cyclopentane from 193 to 298 K and Pressures up to 104 MPa. J Phys Chem 1992;96:1859–68.
[24] Peng D-Y, Robinson DB. A New Two-Constant Equation of State. Ind Eng Chem Fundam 1976;15:59–64.
[25] Poling BE, Prausnitz JM, O'Connell JP. The Estimation of Physical Properties. Prop. Gases Liq., New York: McGraw-Hill; 2004, p. 1–9.
[26] Daubert TE. Vapor - Liquid Critical Properties of Elements and Compounds . 5 . Branched Alkanes and Cycloalkanes. J Chem Eng Data 1996;41:365–72.
[27] Ambrose D, Patel NC. The correlation pressures and estimation of vapour IV. Extrapolation of vapour pressures and estimation of critical pressures the principle of corresponding states using two reference fluids with non-spherical molecules. J Chem Thermodyn 1984;16:459–68.
[28] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.
[29] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic Rankine cycles. Energy 2014;73:204–13.
[30] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.
[31] Iman R, Conover W. A distribution-free approach to inducing rank correlation among input variables. Commun Stat Part B-Simulation Comput 1982;11:311–34.

Paper F

Working fluid selection - Impact of uncertainty of fluid properties

Jérôme Frutiger[a], Jesper Andreasen[b], Wei Liu[c], Hartmut Spliethoff[c], Fredrik Haglind[b], Jens Abildskov[a], Gürkan Sin[a]

a) Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
b) Department of Mechanical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
c) Institute for Energy Systems, Technische Universität München, Garching, Germany

published in Energy (ISSN: 0360-5442), 109, pages: 987-997, 2016.
DOI: http://dx.doi.org/10.1016/j.energy.2016.05.010

Abstract

This study presents a generic methodology to select working fluids for Organic Rankine Cycles (ORC) taking into account property uncertainties of the working fluids. A Monte Carlo procedure is described as a tool to propagate the influence of the input uncertainty of the fluid parameters on the ORC model output, and provides the 95%-confidence interval of the net power output with respect to the fluid property uncertainties. The methodology has been applied to a molecular design problem for an ORC using a low-temperature heat source and consisted of the following four parts: 1) formulation of process models and constraints 2) selection of property models, i.e. Peng-Robinson equation of state 3) screening of 1965 possible working fluid candidates including identification of optimal process parameters based on Monte Carlo sampling 4) propagating uncertainty of fluid parameters to the ORC net power output. The net power outputs of all the feasible working fluids were ranked including their uncertainties. The method could propagate and quantify the input property uncertainty of the fluid property parameters to the ORC model, giving an additional dimension to the fluid selection process. In the given analysis 15 fluids had an improved performance compared to the base case working fluid.

1. Introduction

In recent years the focus on utilization of low-temperature heat sources in different applications such as waste heat in chemical industries and refrigeration plants as well as renewable energy sources such as biomass combustion, geothermal and solar heat sources has increased. The ORC power plant is an important technology to convert this heat into usable work, because it can be applied to a variety of heat sources and a wide range of temperatures [1]. When optimizing the performance of the ORC system it is vital to consider the influence of the working fluid, the component design and the operating conditions. In the early design stage multi-criteria database search and Computer Aided Molecular Design (CAMD) can be applied to generate, test and evaluate promising pure component and mixture candidates as working fluids to help optimize cycle design and performance. Generally speaking CAMD tries to identify the best suitable molecule subject to desired target properties of a defined system.

Database searches and other CAMD methodologies rely crucially on experimental and predicted property data. In the preliminary design stage, a large amount of property data is usually either screened or generated and tested. However, these data are subject to uncertainty, e.g. caused by the measurements [2] or by the property prediction models [3]. In particular the widely used group contribution (GC) models can be subject to varying uncertainty depending on the compound [4]. In the scope of good modeling practices (GMoP), it is necessary to take these property uncertainties into account in order to establish the application range and the reliability of the overall design model [5]. However, there is still a lack of uncertainty analysis methods due to property uncertainty in complex molecular design problems, especially in the domain of working fluid selection and design [6]. Maranas [7] described optimal molecular design considering uncertainty of nonlinear structure-property functionalities. Kim et al. [8] studied the selection of solvents for chemical downstream processes (i.e. extraction) considering uncertainties associated with property estimation. Martín et al. [9] addressed uncertainties due to external factors (e.g. product price) and internal factors in the design of formulated products. Recently the effects of property prediction uncertainty in product design has been considered via the approach of fuzzy optimization [10][11].

The reviews on fluid selection provided by Bao et al. [12] and fluid design studies by Linke et al. [13] give a broad overview of the abundant literature, which is available on working fluids

for ORCs. Recently the combination of fluid design and selection with cycle process optimization has become established as a promising approach to achieve high net power outputs for ORCs with low-temperature heat sources [14]. Chys et al. [15] optimized the process parameters of a large number of working fluids (pure fluids, binary mixtures and three-component mixtures) in low-temperature ORCs to reach maximum thermal efficiency. Andreasen et al. [16] performed a fluid selection and optimization study of ORCs from a low-temperature heat source, considering a large group of binary mixtures as possible working fluids, combined with an evaluation of parameters which affect the design of components. Luo et al. [17] evaluated working fluids with low Global Warming Potential (GWP) in the context of different resource temperatures. Based on a variety of technical, economic and safety-related criteria Rödder et al. [18] selected fluids for a two-stage cycle consisting of a high and low temperature part. Chys et al. [15], Andreasen et al. [16], Luo et al. [17], Rödder et al. [18] used the well-established REFPROP database [19] as source for thermodynamic property data. REFPROP provides uncertainties on measurable quantities like heat capacities, speed of sound and densities. The uncertainties are specific for each fluid and based on the reference where the corresponding data is from. However, there is no information on propagated uncertainty on the corresponding enthalpy or entropy output.

Brown et al. [20] applied the Peng-Robinson equation of state (EoS) to screen the performance potentials of many thousands of working fluids in ORCs. The same authors also varied the fluid parameters of the Peng-Robinson EoS in order to investigate theoretically ideal working fluids [21]. Predicted and experimental property data were used, without propagating its uncertainty to the model output. Stijepovic et al. [22] explored the relationship between working fluid properties and economic and thermodynamic ORC performance criteria. Huixing et al. [23] analyzed the influence of working fluid properties on the ORC cycle performance by optimizing their system for a variety of hydro carbons and hydro fluorocarbons. Desideri et al. [24], Hærvig et al. [25] and Xu et al. [26] studied the influence of the critical temperature of the working fluid on the ORC performance.

Furthermore, there is a variety of studies [27][28][29] that screened working fluid candidates using property data from the well-established DIPPR 801 AIChE database [30] for ORCs with low and high temperature heat sources. For example Drescher and Brüggemann [27] used the DIPPR database for ORC fluid selection in low-temperature biomass power and heat plants. However, none of these studies integrated the data uncertainty information, which is provided by the database, into the modeling.

Papadopoulos et al. [31] used CAMD to optimize the molecular structure of pure components as well as the composition of mixtures and subsequently evaluated the optimum molecules in an ORC process by multi-objective optimization. In a further study Papadopoulos et al. [32] applied CAMD for the synthesis and selection of binary working fluid mixtures for ORC power plants and included a nonlinear sensitivity analysis method to address model-related uncertainties in the mixture selection procedure. The sensitivity analysis allowed identifying and quantifying the model parameters that mainly influence the performance of the mixture candidates in the ORC[32]. Palma-Flores et al. [33] used combined CAMD and process equations in a mixed integer non-linear programming (MINLP) model. Molina-Thierry [34] simultaneously optimized fluid mixtures (generated from a pre-specified set of pure fluids) and the operating conditions of an ORC. The latter three studies used Group Contributions (GCs) and standard cubic EoS in order to estimate the respective fluid properties. The sensitivity analysis method of Papadopoulos et al. [32] was specifically adapted by Mavrou et al. [35] for the identification of optimal fluid mixtures under changing design and operating parameters. Lampe et al. [36] suggested an optimization-based method for the design of optimum ORC working fluids, namely the continuous molecular targeting (CoMT-CAMD) method, which

uses perturbed chain statistical associating fluid theory (PC-SAFT) EoS and quantitative structure-property relationships (QSPR) to estimate properties.

The influence of property uncertainty on the output of a complex molecular design problem, such as the identification of suitable working fluid candidates for ORC power plants, is, to our best knowledge, not yet established. All of the above mentioned studies use a variety of property models, only a few [32][35] consider explicitly the overall influence of their uncertainty on the model output. In this context we believe that the following aspects have not drawn sufficient attention in the literature:

- Implementation of property-focused uncertainty analysis into the working fluid design problem
- Usage of computationally efficient stochastic methods such as Monte Carlo procedure to account for uncertainties in design problems for working fluids
- Inclusion of uncertainty information into the selection process of working fluids

In general there are three types of uncertainties associated with predictions of model simulation: 1) stochastic uncertainty due to stochastic components (e.g. random failure) of a simulation, 2) structural uncertainty related to the approximation of a real physical system by a (generally more simplified) mathematical model, and 3) input uncertainty representing incomplete knowledge about the fixed parameters used as input to the model [5][37]. The most frequently used uncertainty analysis methods in science and engineering are Bayesian analysis and Monte Carlo methods [38][39]. In complex numerical models the use of Bayesian analysis is emerging to perform uncertainty analysis in combination with evolutionary optimization algorithms [40].

In this study a Monte Carlo procedure is applied. Monte Carlo methods are known as well-established method for the propagation of input uncertainty in a variety of applications in science and engineering [41]. The basic principle is to characterize each input parameter of a model by distributions. These distributions are assumed to statistically represent the degree of belief with respect to where the appropriate values of the parameters lie. In order to combine different parameter sets, a Monte Carlo based sampling method using Latin hypercube sampling needs to be applied. Evaluating the model with respect to the parameter sets obtained by sampling from the distributions allows displaying the distribution of the model output, which provides a complete representation of the uncertainty of the model output [41].

A classical attempt is a “one-factor-at-a-time approach” by simply varying the parameters of the model individually while all other parameters remain at their nominal values. This approach however studies the local effect of the parameters in question and therefore the interpretation is valid only locally. However, Monte Carlo based procedures differ in two important ways: (i) the method consider the impact of more than one change of parameter at the time (simultaneous variation of model parameters. Hence the results do not depend on the point where the analysis is carried) and (ii) the method cover a wider range for parameter values (not only variation around nominal values but much larger range specified by the user). Therefore this method depicts the global effect of parameter uncertainty which helps obtain statistically meaningful analysis [41].

The uncertainty analysis methodology follows the work of Sin et al. [5]: Monte Carlo analysis of uncertainty involves three steps: (1) specifying input uncertainty (2) sampling input uncertainty and (3) propagating the sampled input uncertainty in order to obtain a prediction uncertainty for the model output (i.e. the net power output of the ORC power plant) [5].

We present a methodology as well as a pseudo-code for efficient implementation to select working fluids for ORC power plants that takes into account fluid property uncertainty. The current work aims at depicting the influence of the input uncertainty of the fluid parameters on the ORC model output. Only the uncertainty with respect to the fluid properties is considered, whereas the process parameters are kept fixed. The particular focus on the influence of property uncertainty depicts a whole new dimension of the ORC working fluid selection in the context of database search and molecular design.

The methodology consists of a cycle model using the Peng-Robinson EoS. The input property parameters for the working fluid are provided by the DIPPR 801 AIChE database[30] which states the uncertainty of experimental and predicted data. Using a Monte Carlo procedure the uncertainty of the input parameters on the model output (e.g. the net power) is assessed.

The paper is organized as follows: (i) the overall methodology is outlined; (ii) the ORC model including the used EoS and the respective property database is presented (iii) the Monte Carlo procedure used to perform uncertainty analysis and design space exploration is presented (iv) the results of the application of the methodology by screening of all the compounds of the DIPPR 801 database are presented (v) the results are compared to those obtained when property modeling is done using the REFPROP 9.0 database [19].

2. Method and Tools

An overview of the methodology divided in different steps is shown in Table 1.

Table 1. Overview of the methodology.

Step 1	**Formulation of ORC process models and constraints**	
Step 2	**Selection of equation of state and property database**	
Step 3	**Model solution for all compounds**	
	Step 3.1	Sampling in process variable search space
	Step 3.2	Evaluation of model for each process variable sample
	Step 3.3	Identification of optimal process variables
Step 4	**Monte Carlo procedure for uncertainty analysis**	
	Step 4.1	Specification of fluid property input uncertainty
	Step 4.2	Sampling of property search space
	Step 4.3	Evaluation of model for each property sample
	Step 4.4	Computation of 95%-confidence interval of the net power
	Step 4.5	Ranking of the fluids including uncertainty

2.1. Step 1: Formulation of ORC process models and constraints

The ORC process investigated in this study is sketched in Figure 1 and was based on the work of Andreasen et al. [16]. The layout consisted of four main components: a pump, a boiler (preheater, evaporator and superheater), a turbine and a condenser. Heat was provided to the cycle by a hot fluid through the boiler and heat was rejected to a coolant in the condenser. The

working fluid was an organic compound, which was circulated by the pump. The base case fluid was 1,1,1,3,3-Pentafluoropropane (R245fa). At the inlet to the pump the working fluid was in a saturated liquid state at a low pressure. The working fluid was pressurized by the pump and was subsequently directed to the boiler, where it was preheated to the saturated liquid state, evaporated and superheated (superheating was optional). The hot pressurized vapor then entered the turbine where mechanical power was produced. The turbine was connected to a generator which converted the mechanical power to electricity. The cycle was completed by condensation of the low pressure vapor at the turbine outlet.

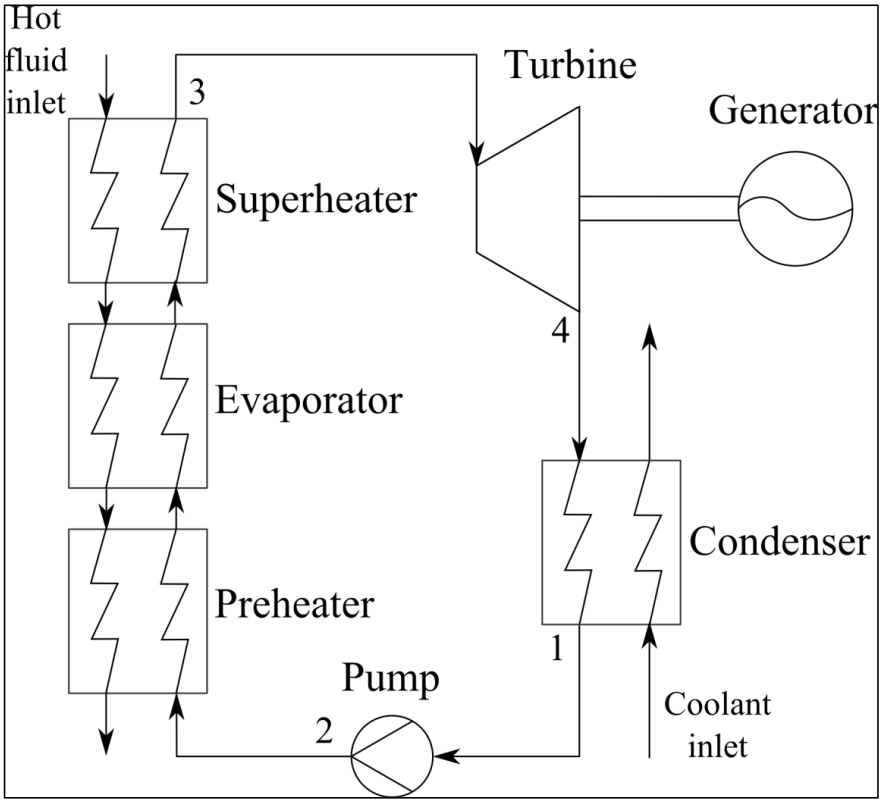


Figure 1. A sketch of the ORC process[16].

The process constraints concerning the conditions used for the hot fluid and the process components are listed in Table 2. The values were adapted from the work of Andreasen et al. [16]. The hot fluid was water at a temperature of 120 °C and a mass flow of 50 kg/s, representative of a waste heat stream of a chemical plant or a geothermal heat source. There were no limitations imposed on the hot fluid outlet temperature. The resulting temperature was therefore the one which ensured thermodynamically optimum conditions for the cycle. Further constraints were: 1) the minimum temperature difference in the boiler was checked at the inlet and outlet (state 2 and 3) and at the saturated liquid point, and 2) the maximum pressure in the boiler was limited at 80 % of the critical pressure of the working fluid. The latter specification ensured that the cycle did not operate too close to the critical point. Hence, this avoided computational problems of the cubic EoS.

Table 2. Modelling constraints for the ORC process [16].

Process parameter	*Value*
Hot fluid inlet temperature	120 °C
Hot fluid mass flow	50 kg/s
Hot fluid pressure	4 bar
Condensation temperature	25 °C
Condenser outlet vapor quality (state 1)	0
Pump isentropic efficiency	0.8
Minimum boiler temperature difference	10 °C
Turbine isentropic efficiency	0.8
Minimum turbine outlet vapor quality (state 4)	1

The assumptions used in the numerical modeling were the following: no pressure losses in piping or heat exchangers, no mechanical or electrical losses, no heat loss from the system, steady state condition and homogeneous flow in terms of thermodynamic properties.

The output from the ORC process model was the net power output, which was the difference between the power production from the turbine and the power consumption of the pump. The net power output $\dot{W}_{NET}$ was calculated from Eq. (1).

$$\dot{W}_{NET} = \dot{m}_{wf}(h_3 - h_4 - (h_2 - h_1)) \quad (1)$$

where *h* is the mass specific enthalpy and $\dot{m}_{wf}$ is the mass flow of the working fluid. The numbering refers to the process overview in Figure 1.

All the process equations are provided in the supporting material.

2.2. Step 2: Selection of equation of state and property database

The thermodynamic properties (i.e. enthalpies and entropies) required in the cycle simulations consisted of an ideal contribution (i.e. the ideal gas enthalpy and entropy) and a correction factor (departure function) accounting for the difference between ideal and real behavior. The Peng-Robinson Equation of State (PR EoS) [42] was selected in order to determine the departure functions of the thermodynamic properties, because of its relatively small number of required fundamental parameters as a cubic equation of state. This made it suitable for the screening of a large number of possible working fluid candidates [43]. Furthermore, PR EoS is generally known to be superior over other standard cubic EoS (e.g. Soave-Redlich-Kwong) for prediction of liquid densities [44].

The Peng-Robinson EOS is given by,

$$p = \frac{R\,T}{V_m - b} - \frac{a\,\alpha}{V_m^2 + 2bV_m - b^2} \quad (2)$$

In Eq. (2) R is the universal gas constant, T is the absolute temperature, p is the pressure, V_m is the molar volume. The other parameters are defined as follows

$$a = \frac{0.457235\,R^2\,T_c^2}{P_c} \quad (3)$$

$$b = \frac{0.077796\, R\, T_c}{P_c} \tag{4}$$

$$\alpha = \left(1 + m\left(1 - T_r^{0.5}\right)\right)^2 \tag{5}$$

$$m = 0.37464 + 1.54226\,\omega - 0.26992\,\omega^2 \tag{6}$$

$$T_r = \frac{T}{T_c} \tag{7}$$

In Eq. (3) to (7) T_c is the critical temperature, P_c is the critical pressure, ω is the acentric factor, T_r is the reduced temperature.

Ideal gas enthalpy and entropy changes were calculated by integrating the temperature-dependent ideal gas heat capacity. The ideal gas heat capacity correlation as defined by Aly and Lee [45] was used. Five compound-specific input parameter (*A*, *B*, *C*, *D*, *E*) (see Eq (8)) were employed,

$$c_p(T) = A + B \cdot \left[\frac{\frac{C}{T}}{sinh(\frac{C}{T})}\right]^2 + D \cdot \left[\frac{\frac{E}{T}}{cosh(\frac{E}{T})}\right]^2 \tag{8}$$

The fluid parameters inputs for Peng-Robinson EoS were the molecular weight *MW*, the critical temperature T_c, critical pressure P_c, and the acentric factor ω. Therefore, the evaluation of the thermodynamic properties required for the ORC model needed only four primary fluid properties (*MW*, T_c, P_c, ω) and the respective Ally-Lee heat capacity constants (*A*, *B*, *C*, *D*, *E*). All these properties could be found in the DIPPR 808 AIChE database [30] for 1965 chemical compounds. The database values for (T_c, P_c, and ω) could be both experimental and predicted. DIPPR provides the Ally-Lee heat capacity constants that had been obtained by fitting the Ally-Lee correlations for each substance to the respective experimental or predicted temperature dependent heat capacity curve. The DIPPR database stated the respective uncertainty of T_c, P_c, and ω along with the heat capacity values obtained from the constants *A*, *B*, *C*, *D*, *E* [46]. This information on the uncertainty was further used to calculate the output uncertainty of the net power.

The detailed property models including all equations are provided in the supporting material.

Given all the process and property equations an analysis of the degrees of freedom could be performed (see Table 3).

Table 3. Degrees of freedom analysis of the combined process and property models

Number of variables	51
Number of equations	33
Number of specifications	8
Degrees of freedom (DOF)	10

There are 10 degrees of freedoms that needed to be fulfilled in order to solve the model. Two degrees of freedom were related to process variables:

Turbine inlet pressure P_t
Turbine inlet temperature T_t,

and 8 degrees of freedom were related to the properties of the working fluid:
Critical pressure P_c
Critical temperature T_c
Acentric factor ω
Heat capacity constants according to Ally and Lee A, B, C, D, E.

The goal of the ORC model was to identify the working fluid that provided the highest net power output. Thus, the task was to identify the best parameter set (P_c, T_c, ω, A, B, C, D, E) out of 1965 compound possibilities together with corresponding optimal process parameters (P_t and T_t) that achieved the highest net power output. In that sense, the problem integrated product and process design aspects.

2.3. Step 3: Model solution for all compounds

We suggest the use of a sampling based approach as means to explore and identify the optimal process variables: For each of the 1965 chemical compounds (defined by the parameters *MW*, T_c, P_c, ω and *A*, *B*, *C*, *D*, *E*) from the DIPPR 808 AIChE database, the optimal process variables (turbine input pressure P_t and temperature T_t) needed to be identified efficiently and the corresponding net power output needed to be calculated. The motivation for using Monte Carlo based sampling approach is to allow fast exploration of process design space for each working fluid candidate , which we have used as an alternative to classical optimization algorithms (e.g. particle swarm optimization)

Step 3.1: Sampling in process variable search space. The Latin hypercube sampling procedure[41] was utilized in order to obtain a number of 250 uniformly distributed pairs of process variables P_t and T_t. The values were sampled within the predefined variable constraints in a temperature range between 25 and 110 °C (corresponding to the condensation temperature and the heat source temperature) and a pressure range between 1 and 15 bars. The lower bound was selected according to the minimum feasible pressure and temperature for the process, whereas the upper bound was fixed in order avoid high safety hazards and costs [27]. Sub-atmospheric pressure was avoided, because it might result in air infiltration, which was undesirable. Furthermore, for low temperature applications the pressures for the optimum fluids were found to be typically above atmospheric pressure [16]. The procedure of Latin hypercube sampling was used to divide the range of each variable in a certain number of equally proportioned intervals. For two variables a two-dimensional square grid was obtained. The samples were positioned such that only one sample exists in each row and each column of the grid (principle of Latin square). The generalization of this concept to more than two variables is the Latin hypercube (Step 4.2) [41].

Step 3.2: Evaluation of process variable samples. For each all of the compounds the ORC model was evaluated using the sampled pairs of process variables. Compounds that did not satisfy constraints inside the model according to Step 1 were screened out within the model evaluation. Subsequently the net power outputs of the corresponding process variable samples were ranked.

Step 3.3: Identification of optimal process variables. At this stage, the process variables giving the highest net power output were chosen to be the most favorable quasi-optimal pair of process variables (P_t and T_t) for the respective compound.

2.4. Step 4: Monte Carlo procedure for uncertainty analysis

For each of the feasible chemical compounds obtained from Step 3 a Monte Carlo based uncertainty analysis focusing on the input property uncertainty was performed.

Step 4.1: Specification of fluid property input uncertainty. The input uncertainties for the parameters T_c, P_c and ω were obtained from DIPPR 801 AIChE database [30]. The uncertainty provided an estimate for the standard deviation of each data point irrespective of whether it had been obtained by experiment or prediction models. For the heat capacity constants (*A, B, C, D, E*) the uncertainty of the calculated heat capacity at a certain temperature using the Ally-Lee correlation was provided. Since the uncertainties of the heat capacity constants themselves were not provided, the overall uncertainty of the calculated heat capacity needs to be taken into account in the model. Therefore, a dummy variable was introduced, which multiplied with the respective ideal gas heat capacity values. The dummy variable *d* had the expectation value 1 and the standard deviation equal to the respective uncertainty. The input parameter space of *MW*, T_c, P_c, ω and *d* was assumed to be normally distributed with a standard deviation equal to the uncertainty and centred around the respective database value. Table 4 showed the input property uncertainty of two selected compounds.

Table 4. Input property uncertainty of two selected compounds

	Property uncertainty				Average relative error
	MW [*g/mol*]	T_c [K]	P_c [bar]	ω [-]	$c_p(T)$ [%]
1,1,1,3,3-Pentafluoropropane	134.0 ± 1.0	427.2 ± 0.9	36.4 ± 3.7	0.380 ± 0.04	3
Perfluoro-n-hexane	338.0 ± 1.0	499.6 ± 5.0	18.0 ± 0.9	0.50 ± 0.01	10
References	[47]	[48][49]	[50][30]	[30]	[51][45]

Step 4.2: Sampling of property search space. The key step of the Monte Carlo procedure is the sampling of the parameter sets. The Latin Hypercube Sampling method was utilized for probabilistic sampling of the fluid property input space of each compound. From the input parameter space a total of 400 samples were selected, each sample containing one value for each input parameter. The sampling range was specified by the uncertainty (i.e. 95%-confidence interval) range of each parameter given by the DIPPR database. The probability of uncertainty is assumed to follow normal distribution for the fluid properties, in contrast to step 3.1, where it was assumed to be uniform for the process variables. The rank-based method for correlation control of Iman and Conover [52] allowed to take into account correlations between the input parameters. This was necessary, because the Peng Robinson EoS parameters were not completely independent. The correlation matrix was directly obtained from the DIPPR data by calculating the respective correlation coefficient between the data sets. From the DIPPR data base values a correlation coefficient was statistically calculated (e.g. between P_c and T_c). Such correlation coefficients are used as input in correlation-based Latin hypercube-sampling procedure. The sampling results were provided in the supporting materials. Figure 2 provided an illustration of the sampling results for the compound 1,1,1,3,3-Pentafluoropropane (R245fa).

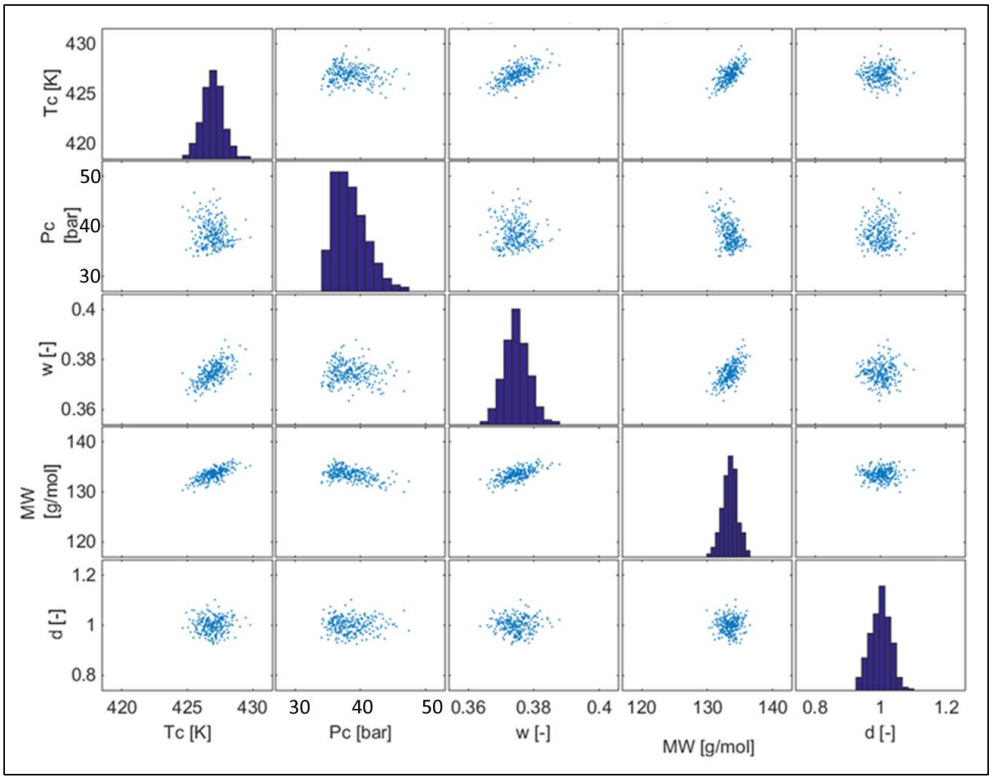


Figure 2. Illustration of samples generated by Latin Hypercube Sampling method with Iman and Conover correlation control for 1,1,1,3,3-Pentafluoropropane (R245fa).

As one can see in Figure 2, the different parameters were assumed to be normally distributed around the property value given in the database, and the width of the distribution was defined by the respective property uncertainty provided from DIPPR. The sampling plots were not uniform, because the correlation between the parameters was taken into account [52].

Step 4.3: Model evaluation for each property sample. One model simulation was performed for each of the 400 input parameter samples resulting from Step 4.2. The Monte Carlo results provided a cumulative distribution function for the net power output of each compound. The uncertainty of the model output could be represented in a pressure-enthalpy- and temperature-entropy-diagram.

Step 4.4: 95%-confidence interval of the net power. Using simple statistics such as mean and percentile calculations, the 95%-confidence interval of the net power output with respect to the corresponding input property values could be obtained for each of the compounds.

Step 4.5: Ranking of the fluids including uncertainty. The compounds were ranked according to their respective net power output including the 95%-confidence interval. This enabled an assessment of the compounds not solely based on their actual cycle performance, but also according to the reliability of the property data used.

The results were compared to the results obtained with REFPROP 9.0 [19] for those compounds for which REFPROP parameters were known to us.

2.5. Modeling platform

The ORC system and the uncertainty analysis were implemented in Matlab (Mathworks, R14) [53]. The methods for performing the uncertainty analysis can be provided upon request as m-script files. Figure 3 depicts the overall methodology as pseudocode.

```
Given N compounds from database
for compound 1:N
    1) Import property data from DIPPR database
    2) Check process constraints
        if feasibility constraints not fulfilled
        Remove compound
        end
    3) Sample S process variables (P,T) pairs by Latin hypercube sampling(uniform distribution)
       without correlation
    4) for sample 1:S
          Evaluate ORC model
          Store net power output for every pair of process variables
       end
    5) Rank Net power outputs
    6) Select process variable pair (P*,T*) with highest net power output as optimal
end
Given M feasible compounds at optimal process variables (T*,P*)
for feasible compound 1:M
   1) Import property input uncertainty from DIPPR database
   2) Sample P property values by Latin hypercube sampling normal distribution with correlation
   3) for sample 1:P
         Evaluate ORC model
         Store net power output for every set of property variables
      end
   4) Depict net power outputs in cumulative distribution function
   5) Calculate mean and standard deviation and 95%-confidence interval
end
Rank Net power output of feasible compounds
```

Figure 3. Pseudocode description of the overall methodology.

3. Results and Discussion

The results are structured in two parts. First the results of the uncertainty analysis are illustrated by the example of one compound, the well-established working fluid 1,1,1,3,3-Pentafluoropropane (R245fa). Afterwards the results of all compounds are analysed and compared.

3.1. Uncertainties with respect to the properties of R245fa

Uncertainty analysis has been performed for all feasible compounds from Step 3.2 (see next section). As an example, the uncertainty analysis results for 1,1,1,3,3-Pentafluoropropane (R245fa) is given in detail. The Monte Carlo simulations obtained by simulating the 400 Latin Hypercube Samples of the property parameters resulted in 260 feasible model solutions for R245fa. The raw data obtained from the simulations can be plotted in a temperature-entropy and a logarithmic pressure-enthalpy diagram showing cycle points enumerated as in Figures 4 and 5. Each curve and design point set is different as different property parameter samples were used in each simulation. A varying band for both the saturation curves and the cycle design can be observed. From a statistical point of view the bands correspond to the

distribution of the model outputs and directly show the sensitivity with respect to the fluid property values for the specific compound R245fa. The larger the width of the output range, the higher the uncertainty. Furthermore, the uncertainty in the model outputs can be represented using mean (solid line) and 95%-confidence interval (dashed line) obtained from percentile calculations.

Since the condensation temperature of the ORC is fixed, there is no variation in the Ts-diagram in the condensation process. Similarly, there is no variation in the evaporation pressure in the Ph-diagram, because the input pressure to the turbine (identified in step 3 of the methodology) is kept fixed. Although these process variables can be subject to uncertainty, this study focused particularly on the property parameter uncertainty.

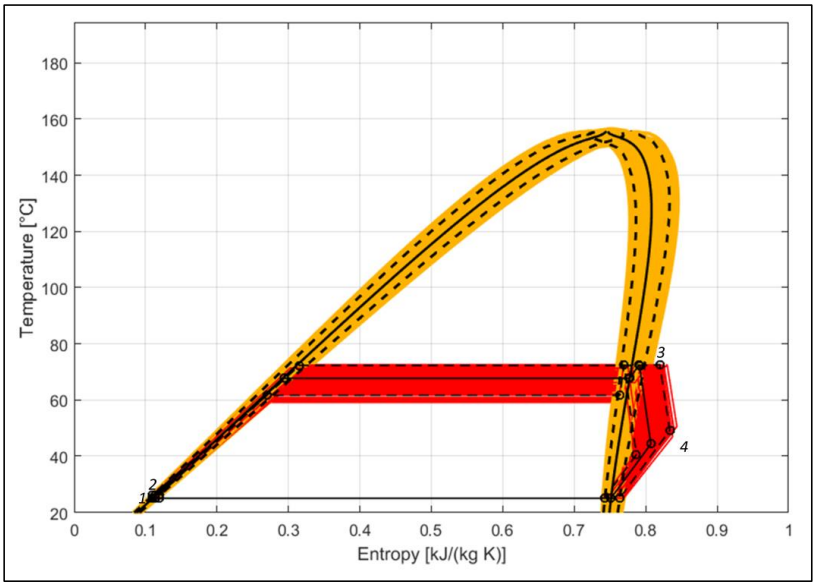


Figure 4. Representation of uncertainty with respect to the fluid properties in the Ts-diagram for R245fa: Monte Carlo simulations (yellow/red), mean (solid line) and upper as well as lower bound of the 95%-confidence interval of the model output (dashed lines).

For each of the simulations of R245fa the net power output can be obtained. The distribution of the net power outputs for R245fa and the corresponding empirical cumulative distribution function (empirical CDF) are two alternative representations of the uncertainty of the model with respect to the property parameters (see Figures 6 and 7). Figure 6 shows a large variety for the net power output due to property uncertainty. The gaps in Figure 6 occur, because some combinations of parameters are infeasible with respect to the property model evaluation.

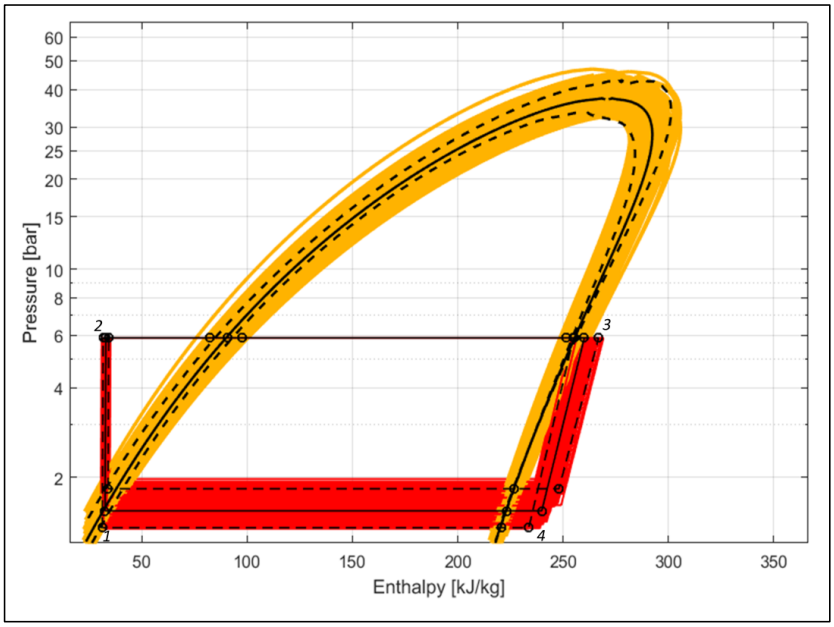


Figure 5. Representation of uncertainty with respect to the fluid properties in the Ph-diagram for R245fa: Monte Carlo simulations (yellow/red), mean (continuous line) and upper as well as lower bound of the 95%-confidence interval of the model output (dashed lines).

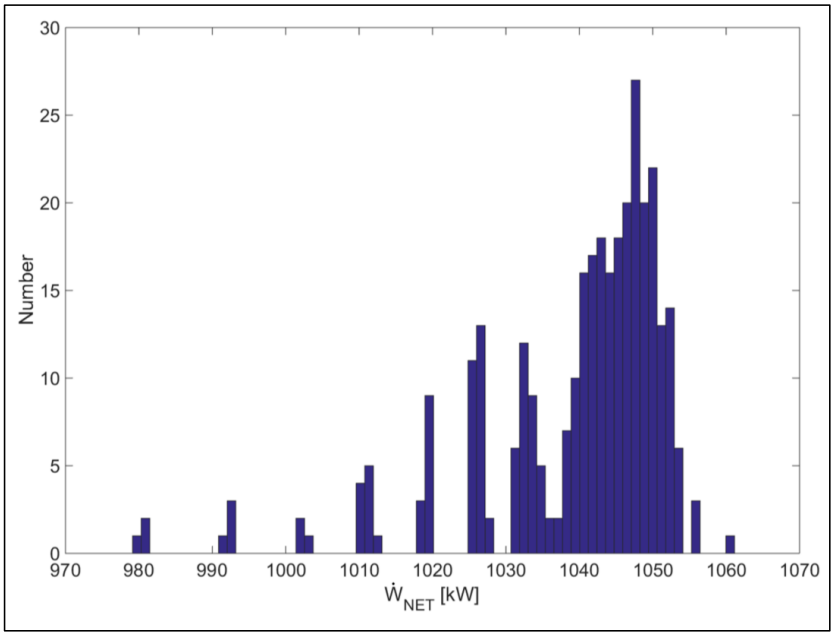


Figure 6. Distribution of the net power output $\dot{W}_{NET}$ of the ORC for R245fa.

The empirical CDF depicted in Figure 7 is a step function that increases by 1/n in every data point, where n is the number of data points. Its aim is to estimate the true underlying

distribution function. It does not a priori assume a normal distribution (or any other distribution function). The empirical CDF allows calculating 2.5% and 97.5% percentile defining the lower and upper bound of the 95%-confidence for the net power output.

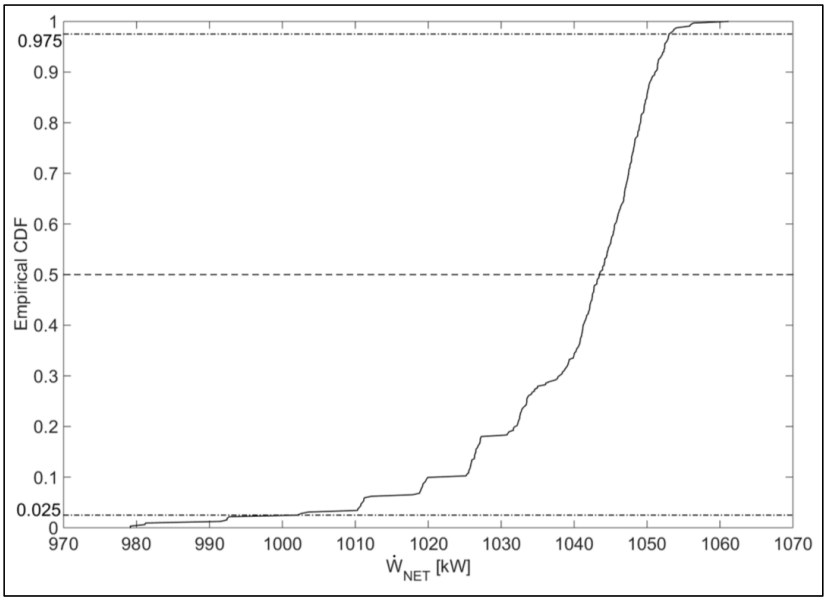


Figure 7. Empirical Cumulative Distribution Function (empirical CDF) of the net power output $\dot{W}_{NET}$ of the ORC for R245fa with mean and 95%-confidence interval of the model output (vertical dotted lines).

As shown in Figures 4, 5, 6 and 7, even small variations of input property parameters (Table 4) can lead to significant output uncertainty. This can be explained by the high non-linearity of the property models and process equations. The model is sensitive to different combinations of property parameters, although the importance of the individual parameters might appear to be small. It is a clear indication that the fluid property uncertainty cannot be neglected for cubic equations of state in the design of ORC, as it has been done in many previous studies and applications. This conclusion is of particular importance for CAMD problems.

Although the study focused on the ORC process, the uncertainty analysis procedure with respect to the input parameter uncertainty can also be performed for other types of thermodynamic cycles, e.g. heat pumps.

3.2. Ranking of working fluids including uncertainty

Having obtained mean and 95%-confidence interval of the net power output of all the fluids, it is possible to rank the compounds. Table 5 shows the $\dot{W}_{NET}$ mean value of the distribution for 20 best performing compounds for the given ORC power plant including their corresponding uncertainty with respect to the property input. Furthermore, the net power outputs are compared to the values obtained by using the corresponding REFPROP correlations. The results of all feasible compounds are given in the supporting material. Furthermore, the ranking does not include safety and environmental properties of the fluid, because the particular focus of this study lies in the analysis of property uncertainty. However, it has to be acknowledged in the current case that the fluoro-compounds have a relatively high Global Warming Potential (GWP> 150) and that the hydrocarbons have a high lower flammability

limit (LFL>0.1 kg/m^3) [54]. The mean value of $\dot{W}_{NET}$ and the corresponding 95%-confidence interval of the compound in Table 4 are represented in Figure 8.

Table 5. Best performing compounds ranked by net power output including uncertainty.

Rank	Working fluid name	Mean $\dot{W}_{NET}$ [kW]	95%-conf. int. $\dot{W}_{NET}$ [kW]		Optimal $\dot{W}_{NET}$ [kW] from Step 3	Rel. deviation to REFPROP 9.0 [%]
1	Decafluorobutane (R610)	1145.8	1188.6	1042.6	1149.2	0.56
2	Octafluorocyclobutane (RC318)	1145.3	1174.6	1110.1	1161.7	0.76
3	2-Chloro-1,1,1,2-Tetrafluoroethane (R124)	1084.1	1096.7	1063.3	1083.7	1.01
4	Perfluoro-n-Pentane	1068.3	1095.4	1016.3	1083.6	1.74
5	Isobutane	1067.4	1074.2	1060.0	1068.4	0.85
6	Pentafluoroethylmethylether	1063.0	1083.5	1038.1	1069.3	n/a
7	Neopentane	1050.8	1056.9	1043.3	1050.3	0.22
8	n-Butane (R600)	1050.1	1055.7	1044.6	1049.6	0.44
9	1-Chloro-1,1-difluoroethane (R142b)	1049.1	1057.1	1031.9	1051.9	0.51
10	Isobutene	1045.7	1051.6	1039.7	1048.2	0.32
11	Trifluoroiodomethane	1043.3	1063.7	151.9	1052.1	1.08
12	1-Butene	1042.7	1048.6	1035.9	1039.5	0.03
13	1,1,1,2,3,3-Hexafluoropropane (R236ea)	1041.7	1059.1	1008.4	1051.4	0.01
14	1,1,1,2,3,3,3-Heptafluoropropane (R227ea)	1040.2	1199.0	645.4	1162.5	8.38
15	Heptafluoropropylmethylether	1039.7	1046.9	1034.0	1040.5	0.41
16	*1,1,1,3,3-Pentafluoropropane (R245fa) – Base case*	*1039.3*	*1054.5*	*1002.5*	*1050.7*	*0.71*
17	trans-2-Butene	1036.9	1041.5	1031.5	1041.2	0.34
18	1,2-Dichlorotetrafluoroethane (R114)	1036.2	1049.4	1010.1	1047.3	0.23
19	Bis(Difluoromethyl)ether	1035.9	1060.7	992.6	1048.5	n/a
20	1,1,1,3,3,3-Hexafluoropropane (R236fa)	1035.8	1111.1	762.5	1106.9	5.10

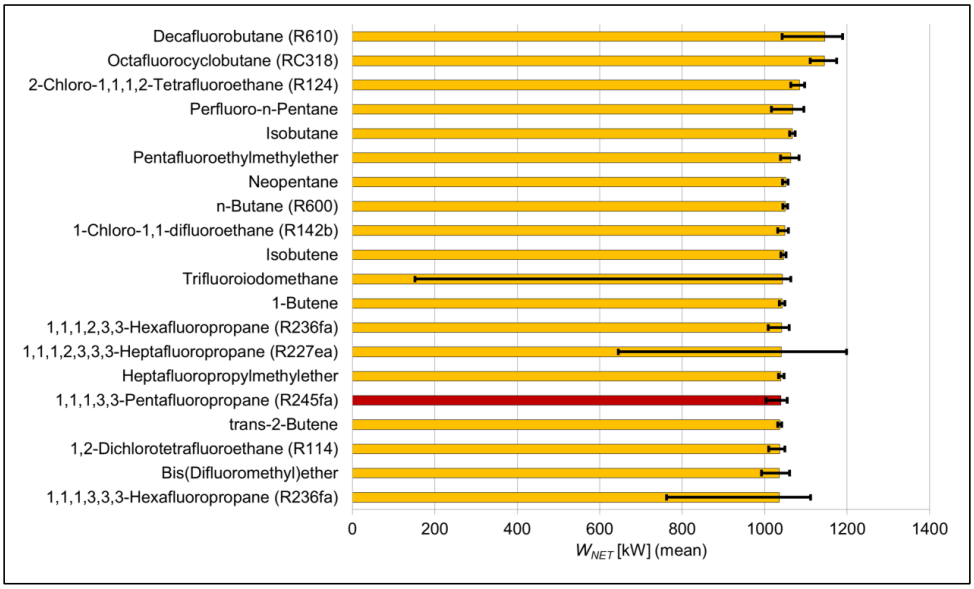


Figure 8. Mean value of the net power output $\dot{W}_{NET}$ of the 20 best performing compounds including the 95%-confidence interval (thin black bars) obtained from the uncertainty analysis with respect to the fluid properties. The red bar corresponds to the base case compound.

As it can be seen in Table 5, the values obtained with REFPROP have (with two exceptions) a relative deviation below 2%, suggesting that the numerical models presented in this paper give reasonable results. The exceptions are 1,1,1,2,3,3,3-Heptafluoropropane (R227ea) with a deviation of 8.64% and 1,1,1,3,3,3-Hexafluoropropane (R236fa) with 5.10%. Both compounds also have a wide range of uncertainty. The deviation with REFPROP and the range of uncertainty suggest that the property data for these compounds need to be measured more accurately and reliably.

The fluids are ranked by the mean value obtained from the distribution. These values are not identical to the optimal net power output obtained in Step 3 of the methodology before the property uncertainty analysis, although the optimal value lies within the respective 95%-confidence interval. The discrepancy is a direct consequence of the fluid property uncertainty analysis. Ranking the compounds only by the optimal net power output without considering uncertainty, would be in particular give a different position for 1,1,1,2,3,3,3-Heptafluoropropane (R227ea) and 1,1,1,3,3,3-Hexafluoropropane (R236fa). Both compounds have a large optimal net power output obtained in Step 3, but due to uncertainty propagation the mean value of the net power output distribution is lower compared to the other compounds.

Alternatively the fluids can be ranked according to their respective lower bound of the 95%-confidence interval (see Figure 9). This is a conservative approach of ranking and can be considered as the statistically robust way to identify promising working fluid candidates.

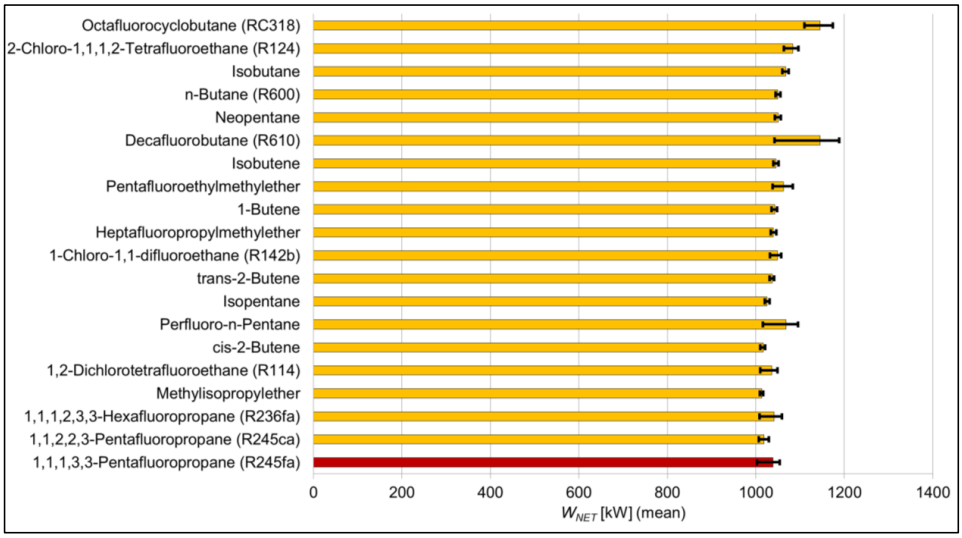


Figure 9. Ranking according to the lower bound value of the 95%-confidence interval of the net power output $\dot{W}_{NET}$. The red bar corresponds to the base case compound.

Knowing the 95%-confidence interval of the net power output for the screened chemical compounds gives an important new dimension in the preliminary selection of suitable working fluid candidates. Some working fluids, e.g. isobutane (Compound 5 in Table 5), have a very small model output uncertainty range, whereas others, e.g. trifluoroiodomethane (Compound 11), have a very large. If the 95%-confidence interval of a compound overlaps with the one of the base case, it is statistically impossible to say, which of two perform better. This is the case for one of the two top compounds. Although the $\dot{W}_{NET}$ mean value of decafluorobutane (Compound 1) and octafluorocyclobutane (Compound 2) are very close to each other, the 95%-confidence interval of decafluorobutane (Compound 1) is overlapping with the 95%-confidence interval of the base case (see Figure 8). This can also be seen in Figure 9, where the ranking was made according to the lower bound of the 95%-confidence interval, as decafluorobutane is not anymore at the top. Hence, the uncertainty analysis provides important additional information for the interpretation of the results. Based on the analysis of this study, the best performing compounds with the smallest uncertainty range is in fact octafluorocyclobutane (Compound 2). However, the study also implies that more reliable property data for decafluorobutane (Compound 1) is needed.

There are two major causes for large net power output uncertainty:

(1) The input property uncertainty of one or more parameters is high thus resulting in a large net power output uncertainty. This is directly related to the reliability of the measured and predicted property data.

(2) The cycle is operated in a sensitive region in terms of the fluid properties with respect to the model evaluation for a particular fluid. Hence, small variations of the parameters have a large impact on the model output. The knowledge of whether a fluid is sensitive to the ORC model structure or not is a priori unknown. Therefore, an uncertainty analysis of the model output with respect to the fluid properties can give vital information. However, only a global sensitivity analysis of the property parameters with respect to the net power output can provide an in-depth investigation of the overall influence of a particular fluid parameter value to the model output, which is beyond the scope of this study.

The range of uncertainty can be considered as a novel criterion in model based working fluid selection. The narrower the 95%-confidence interval, the more reliable the property data and the less sensitive the fluid performs in the cycle. This information is vital for further detailed modeling and experimental validation studies of identified promising fluids.

4. Conclusion

The study presented a methodology to select working fluids for ORCs considering uncertainty of fluid property estimations. The uncertainty values were taken from DIPPR AIChE database including both experimental measurements and property prediction methods (e.g. GC methods). The fluid property uncertainty was propagated using the Monte Carlo procedure to estimate the net power output uncertainty. Furthermore, a large amount of compounds from the DIPPR database were screened and subsequently the uncertainty analysis method was applied.

The following are the main conclusions from the systematic screening and the uncertainty analysis:

- The uncertainty analysis with respect to the input property uncertainty is a vital tool for ORC model analysis.
- The Monte Carlo based procedure can be applied to propagate fluid property uncertainty to the model output, independent of the process and property models.
- Calculating the net power output including its 95%-confidence interval for each fluid, gives an additional quantitative criterion for the fluid selection assessing fluid data uncertainty and model sensitivity.
- The ranking of working fluids can be significantly different based whether the mean value or uncertainties (e.g. the lower bound of the 95%-confidence interval) of the net power output are used.
- In this study the screening of working fluid candidates identified octafluorocyclobutane as the best performing working fluid with smallest model output uncertainty.

We suggest that future ORC working fluid studies should take into account fluid property uncertainty as a tool to base any kind of fluid investigation, comparison or selection on a thorough property and process model analysis.

References

[1] Invernizzi CM. The Organic Rankine Cycle. Closed Power Cycles, Thermodyn. Fundam. Appl., London: Springer Verlang; 2013, p. 117–76.

[2] Dong Q, Chirico RD, Yan X, Hong X, Frenkel M. Uncertainty Reporting for Experimental Thermodynamic Properties. J Chem Eng Data 2005;50:546–50.

[3] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.

[4] Frutiger J, Marcarie C, Abildskov J, Sin G. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models – an application to heat of combustion. J Chem Eng Data 2016;61:602–13.

[5] Sin G, Gernaey K V., Eliasson Lantz A. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. Biotechnol Prog 2009;25:1043–53.

[6] Ng LY, Chong FK, Chemmangattuvalappil NG. Challenges and opportunities in computer-aided molecular design. Comput Chem Eng 2014;81:115–29. doi:10.1016/j.compchemeng.2015.03.009.

[7] Maranas C. Optimal molecular design under property prediction uncertainty. AICHE J 1997;43:1250–64.

[8] Kim K-J, Diwekar UM. Efficient Combinatorial Optimization under Uncertainty. 1. Algorithmic Development. Ind Eng Chem Res 2002;41:1276–84.

[9] Martín M, Martínez A. Addressing Uncertainty in Formulated Products and Process Design. Ind Eng Chem Res 2015;54:5990–6001.

[10] Ng LY, Chemmangattuvalappil NG, Ng DKS. Robust chemical product design via fuzzy optimisation approach. Comput Aided Chem Eng 2014;34:387–92.

[11] Ten JY, Hassim MH, Chemmangattuvalappil N, Ng DKS. A Novel Chemical Product Design Framework with the Integration of Safety and Health Aspects. J Loss Prev Process Ind 2015;40:67–80.

[12] Bao J, Zhao L. A review of working fluid and expander selections for organic Rankine cycle. Renew Sustain Energy Rev 2013;24:325–42.

[13] Linke P, Papadopoulos A, Seferlis P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. Energies 2015;8:4755–801.

[14] Larsen U, Pierobon L, Haglind F, Gabrielii C. Design and optimisation of organic Rankine cycles for waste heat recovery in marine applications using the principles of natural selection. Energy 2013;55:803–12.

[15] Chys M, Broek M Van Den, Vanslambrouck B, Paepe M De. Potential of zeotropic mixtures as working fl uids in organic Rankine cycles. Energy 2012;44:623–32.

[16] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic Rankine cycles. Energy 2014;73:204–13.

[17] Luo D, Mahmoud A, Cogswell F. Evaluation of Low-GWP fluids for power generation with Organic Rankine Cycle. Energy 2015;85:481–8. doi:10.1016/j.energy.2015.03.109.

[18] Rödder M, Neef M, Laux C, Priebe KP. Systematic fluid selection for organic Rankine cycles (ORC) and performance analysis for a combined high and temperature low temperature cycle. Proc ASME Turbo Expo 2015 Turbine Tech Conf Expo 2015;138:1–9.

[19] Lemmon E, Huber M, McLinden M. Reference fluid thermodynamic and transport properties-REFPROP, standard reference database 23, version 8.0, National Institute of Standard and Technology; 2007.

[20] Brown JS, Brignoli R, Daubman S. Methodology for estimating thermodynamic parameters and performance of working fluids for organic Rankine cycles. Energy 2014;73:818–28.

[21] Brown JS, Brignoli R, Quine T. Parametric investigation of working fluids for organic Rankine cycle applications. Appl Therm Eng 2015;90:64–74.

[22] Stijepovic MZ, Linke P, Papadopoulos AI, Grujic AS. On the role of working fluid properties in Organic Rankine Cycle performance. Appl Therm Eng 2012;36:406–13.

[23] Zhai H, Shi L, An Q. Influence of working fluid properties on system performance and screen evaluation indicators for geothermal ORC (organic Rankine cycle) system. Energy 2014;74:2–

11.
[24] Desideri A, Gusev S, Broek M Van Den, Lemort V, Quoilin S. Experimental comparison of organic fluids for low temperature ORC systems for waste heat recovery applications 2015;97:460–9.
[25] Hærvig J, Sørensen K, Condra TJ. Guidelines for optimal selection of working fluid for an organic Rankine cycle in relation to waste heat recovery. Energy 2016;96:592–602.
[26] Xu J, Yu C. Critical temperature criterion for selection of working fluids for subcritical pressure Organic Rankine cycles. Energy 2014;74:719–33.
[27] Drescher U, Brüggemann D. Fluid selection for the Organic Rankine Cycle (ORC) in biomass power and heat plants. Appl Therm Eng 2007;27:223–8.
[28] Price SE, Mayor JR. Thermodynamic Design and Comparative Analysis of Rankine, ORC and Kalina Cycles for Low-Cost, Meso-Scale Power Generation Systems. ASME 2008 2nd Int. Conf. Energy Sustain., 2008.
[29] Romanin, Vincent D. Rodriguez A, Jean T, Fereres S. Using the Peng-Robinson Equation of State to Explore Working Fluids for Higher Temperature Organic Rankine Cycles. ASME 2014 Int. Mech. Eng. Congr. Expo. Vol. 6A Energy, 2014.
[30] Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE) 2014.
[31] Papadopoulos AI, Stijepovic M, Linke P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. Appl Therm Eng 2010;30:760–9.
[32] Papadopoulos AI, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward Optimum Working Fluid Mixtures for Organic Rankine Cycles using Molecular Design and Sensitivity Analysis. Ind Eng Chem Res 2013;52:12116–33.
[33] Palma-Flores O, Flores-Tlacuahuac A, Canseco-Melchor G. Optimal molecular design of working fluids for sustainable low-temperature energy recovery. Comput Chem Eng 2015;72:334–49.
[34] Molina-Thierry DP, Flores-Tlacuahuac A. Simultaneous optimal design of organic mixtures and rankine cycles for low-temperature energy recovery. Ind Eng Chem Res 2015;54:3367–83.
[35] Mavrou P, Papadopoulos AI, Seferlis P, Linke P, Voutetakis S. Selection of working fluid mixtures for flexible Organic Rankine Cycles under operating variability through a systematic nonlinear sensitivity analysis approach. Appl Therm Eng 2015;89:1054–67. doi:10.1016/j.applthermaleng.2015.06.017.
[36] Lampe M, Stavrou M, Schilling J, Sauer E, Gross J, Bardow A. Computer-aided Molecular Design in the Continuous-Molecular Targeting Framework using Group-Contribution PC-SAFT. Elsevier Ltd; 2015.
[37] McKay MD, Morrison JD, Upton SC. Evaluating prediction uncertainty in simulation models. Comput Phys Commun 1999;17:44–51.
[38] Helton JC. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. vol. 57. 1997.
[39] Chandrashekar N, Krishnamurty S. Bayesian evaluation of engineering models. ASME 2002 Int Des Eng Tech Conf Comput Inf Eng Conf 2002;2:993–1002.
[40] Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. London: Chapman & Hall/CRC; 2003.
[41] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.
[42] Peng D-Y, Robinson DB. A New Two-Constant Equation of State. Ind Eng Chem Fundam 1976;15:59–64.
[43] Liu W, Meinel D, Wieland C, Spliethoff H. Investigation of hydrofluoroolefins as potential working fluids in organic Rankine cycle for geothermal power generation. Energy 2014;67:106–16.
[44] Pierre D. Essentials of Reservoir Engineering. 2007.
[45] Aly F, Lee L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. Fluid Phase Equilib 1981;6:169–79.
[46] Knotts T, Wilding WV, Rowley R, Giles N, Congore A. The DIPPR 801 Gold Standard Systems Approach to Critical Evaluation of Thermophysical Property Data. 19th Symp. Thermophys. Prop., Boulder: 2015.
[47] IUPAC. Atomic Weights of the Elements 1999. J Phys Chem Ref Data 1999;30:701–12.

[48] Schmidt JW, Carrillo-Nava E, Moldover MR. Partially Halogenated Hydrocarbons CHFCl-CF3, CF3-CH3, CF3-CHF-CHF2, CF3-CH2-CF3, CHF2-CF2-C2F3, CF3-CH2-CHF2, CF3-O-CH2; Critical Temperature, Refractive Indices, Surface Tension and Estimates of Liquid, Vapor and Critical Densities. Fluid Phase Equilib 1996:187–206.
[49] Cheng DC-H, McCoubrey JC. The Critical Temperatures of Covalent Fluorides. J Chem Soc 1963:4993–5.
[50] Sotani T, Kubota H. Vapor Pressures and PVT Properties of 1,1,1,3,3-Pentafluoropropane (HFC-245fa). Fluid Phase Equilib 1999:32–355.
[51] Bures M, Majer V, Zabransky M. Modification of Benson method for estimation of ideal-gas heat capacities. Chem Eng Sci 1981;36:529–37.
[52] Iman R, Conover W. A distribution-free approach to inducing rank correlation among input variables. Commun Stat Part B-Simulation Comput 1982;11:311–34.
[53] MATLAB 2015b, The MathWorks, Inc., Natick, Massachusetts, United States n.d.
[54] Regulation (EC) No 1005/2009 of the European Parliament and of the council. Off J Eur Communities 2000;L 269:1–15.

Paper appendix

A.2. Data sheets

Detailed ORC and property model description. List of all feasible compounds including the calculated net power output and the corresponding uncertainties.

The data sheets are permanently available on the webpage of Energy: http://dx.doi.org/10.1016/j.energy.2016.05.010.

A.2. Nomenclature

Acronymns

GMoP	good modeling practices
CAMD	computer aided molecular design
EoS	equation of state
CoMT-CAMD	continuous molecular targeting computer aided molecular design
PC-SAFT	perturbed chain statistical associating fluid theory
QSPR	quantitative structure-property relationships
ORC	organic Rankine cycle
GWP	global warming potential
LFL	lower flammability limit

Symbols

P	pressure [bar]

R	ideal gas constant [J/(molK)]
T	temperature [C]
V*m*	molar volume [m3/mol]
α	parameter of Peng-Robinson equation of state
m	parameter of Peng-Robinson equation of state
b	parameter of Peng-Robinson equation of state
ω	acentric factor [-]
T_r	reduced temperature [-]
P_c	critical pressure [bar]
T_c	critical temperature [C]
$\dot{W}_{NET}$	net power output [kW]
MW	molecular weight [g/mol]
A	constant of temperature dependent heat capacity [J/molK]
B	constant of temperature dependent heat capacity [J/molK]
C	constant of temperature dependent heat capacity [J/molK]
D	constant of temperature dependent heat capacity [J/molK]
E	constant of temperature dependent heat capacity [J/molK]
$c_p(T)$	temperature dependent heat capacity [J/(mol K)]
P_t	turbine inlet pressure [bar]
T_t	turbine inlet temperature [C]
m_{wf}	working fluid mass flow [kg/s]
CDF	cumulative distribution function
h_i	specific enthalpy at position i [kJ/kg]
d	dummy variable

Subscripts and superscripts

c	critical
i	position in the cycle
t	turbine
wf	working fluid

p	constant pressure
r	reduced
m	molar
NET	net

A.3. Property model description

The property model based on Peng-Robinson Equation of State (PR-EOS) is shown. For the current model PR-EOS is known to be powerful in the prediction of both liquid and vapor properties. The PR-EOS is adapted from the work of Liu et al. (2014) [1].

The given property data of a certain compound are:

- Critical pressure P_c
- Critical temperature T_c
- Accentric factor ω
- Heat capacity constants according to Ally and Lee (1981) [2] A, B, C, D, E

Liu et al. (2014) [1] used the formulation of PR-EOS according to Robinson, Peng and Chung (1985) [3].

The PR-EOS is described according to the original work of Peng and Robinson (1976) [4].

The Peng-Robinson equation (PR-EOS) is defined as in Eq. (9)

$$P = \frac{R\,T}{V_m - b} - \frac{a\,\alpha}{V_m^2 + 2bV_m - b^2} \tag{9}$$

In Eq. (9) R is the universal gas constant, T is the absolute temperature, P is the pressure, V_m is the molar volume. The other parameters are defined as follows

$$a = \frac{0.457235\,R^2\,T_c^2}{p_c} \tag{10}$$

$$b = \frac{0.077796\,R\,T_c}{p_c} \tag{11}$$

$$\alpha = \left(1 + m\left(1 - T_r^{0.5}\right)\right)^2 \tag{12}$$

$$m = 0.37464 + 1.54226\,\omega - 0.26992\,\omega^2 \tag{13}$$

$$T_r = \frac{T}{T_c} \tag{14}$$

In Eq. (10) to (14) T_c is the critical temperature, p_c is the critical pressure, ω is the acentric factor, T_r is the reduced temperature.

In order to avoid convergence problems in the iterative calculation procedure, PR-EOS is rewritten in polynomial form by introducing the compressibility factor Z [5].

$$Z = \frac{pV}{RT} \tag{15}$$

In that sense PR-EOS is formulated as in Eq. (16) [5]

$$Z^3 - (1 - B)\ Z^2 + (A - 2B - 3B^2)\ Z - (AB - B^2 - B^3) = 0 \tag{16}$$

where A and B are given as

$$A = \frac{a\alpha p}{R^2\ T^2} \tag{17}$$

$$B = \frac{bp}{RT} \tag{18}$$

The fugacity coefficient for PR-EOS is defined as follows [5]

$$\phi = exp\left(Z - 1 - ln(Z - B) - \frac{A}{2 \cdot B \cdot \sqrt{2}} \cdot ln\left(\frac{Z + (1 + \sqrt{2}\cdot) \cdot B}{Z + (1 - \sqrt{2}) \cdot B}\right)\right) \tag{19}$$

Equilibrium conditions, such as saturated pressure p_{sat} or boiling point T_{sat} can be found by solving the isofugacity condition between vapor and liquid. This is needed, when solving the PR-EoS in order to determine the saturation conditions.

$$\phi^L = \phi^V \tag{20}$$

A cubic equation of state like PR-EOS has three roots, where physically meaningful roots, i.e. solutions for Z, are real and positive.

- If $T \geq T_c$, the solution of a cubic equation of states gives only one solution. This is also true for $T = T_c$, but in the latter case the solution is the critical compressibility factor Z_c.
- If $T < T_c$ the cubic equation of states may give three real and positive solutions. The smallest root is liquid-like Z_{liquid} and the largest root vapor-like Z_{vapor}. The third root lying between is of no physical significance. There exists three possible cases:
    1. At saturation condition ($P = P_{sat}(T)$ or $T = T_{sat}(P)$): The smallest root is the liquid phase Z_{liquid} and the largest root Z_{vapor} is the vapor phase.
    2. Above saturation condition ($P > P_{sat}$ and $T > T_{sat}$): The largest root Z_{vapor} (vapor-like) needs to be chosen to characterize the fluid in vapor phase.
    3. Below saturation condition ($P < P_{sat}$ and $T < T_{sat}$): The smallest root Z_{liquid} (liquid-like) needs to be chosen to characterize the fluid in liquid phase.

  Hence, the chosen pressure and temperature need to be compared to saturation conditions [6].

After solving PR-EOS the three state properties are known: temperature, pressure, volume. Hence, caloric properties can be calculated.

The enthalpy of a certain state can be calculated by Eq. (21)

$$h(p,T) = h^{ref} + h(T)^{ideal} + h(P,T)^{PR\,dep} \tag{21}$$

where h^{ref} is the enthalpy of the defined reference state at $T = T^{ref}$ and $P = P^{ref}$, $h(T)^{ideal}$ is the enthalpy of the ideal gas as function of temperature and $h(P,T)^{PR\,dep}$ is the departure function from PR-EOS. The departure function describes the difference from the ideal state to the real state.

The change in enthalpy of an ideal gas when temperature changes from T_{ref} to T can be calculated by integrating the temperature-dependent heat capacity $c_p(T)$. The heat capacity correlation is defined according to Aly and Lee (1981) [2].

$$h(T)^{ideal} = \int_{T_{ref}}^{T} c_p(T)\,\mathrm{d}T \tag{22}$$

$$h(T)^{ideal} = \int_{T_{ref}}^{T} A + B \cdot \left[\frac{\frac{C}{T}}{sinh(\frac{C}{T})}\right]^2 + D \cdot \left[\frac{\frac{E}{T}}{cosh(\frac{E}{T})}\right]^2 \mathrm{d}T \tag{23}$$

In Eq. (8) A, B, C, D and E are the heat capacity constants according to Aly and Lee (1981).

Eq. (8) can be evaluated analytically.

$$\begin{aligned} h(T)^{ideal} &= A \cdot (T - T_{ref}) + BC[coth(C/T) - coth(C/T_{ref})] \\ . &\quad - DE[tanh(E/T) - tanh(E/T_{ref})] \end{aligned} \tag{24}$$

The departure function of enthalpy for the Peng-Robinson equation is shown in Eq. (25) [1]

$$h(P,T)^{PR\,dep} = RT_c\left[T_r(Z-1) - 2.078(1+m)\sqrt{\alpha}\ln\left(\tfrac{Z+2.414B}{Z-0.414B}\right)\right] \tag{25}$$

where α is defined as in Eq. (12), *m* is defined as in Eq. (13) and B is given by Eq. (18).

In analogy to the enthalpy the entropy at a certain temperature and pressure can be computed using PR-EOS (see Eq. (26))

$$s(p,T) = S^{ref} + S(T,P)^{ideal} + S(P,T)^{PR\,dep} \tag{26}$$

where s^{ref} is the reference state, $s(T)^{ideal}$ is the entropy of the ideal gas as function of temperature and $s(P,T)^{PR\,dep}$ is the departure function from Peng-Robinson-EOS.

The change in entropy of an ideal gas when temperature changes from T_{ref} to T and pressure changes from P_{ref} to P is defined as follows.

$$s(T,P)^{ideal} = \int_{T_{ref}}^{T} c_p(T)\,\mathrm{d}T - \int_{P_{ref}}^{P} V dP \tag{27}$$

Using the ideal gas equation $pV = RT$, it is possible to substitute V:

$$s(T,P)^{ideal} = \int_{T_{ref}}^{T} \frac{c_p(T)}{T}\,\mathrm{d}T - \int_{P_{ref}}^{P} R \cdot \frac{1}{P} dP \tag{28}$$

Furthermore, it is possible to integrate the temperature-dependent heat capacity $c_p(T)$ according to Aly and Lee (1981) [2] .

$$\begin{aligned} s(T,P)^{ideal} = & \int_{T_{ref}}^{T} \frac{1}{T} \left(A + B \cdot \left[\frac{\frac{C}{T}}{sinh(\frac{C}{T})} \right]^2 + D \cdot \left[\frac{\frac{E}{T}}{cosh(\frac{E}{T})} \right]^2 \right) \mathrm{d}T \\ . & - R \cdot log\left(\frac{P}{P_{ref}} \right) \end{aligned} \tag{29}$$

$$\begin{aligned} s(T,P)^{ideal} = & A \cdot ln(\frac{T,P}{T_{ref}}) + BC \left(\frac{coth(C/T)}{T} - \frac{coth(C/T_{ref})}{T_{ref}} \right) - \\ . & B \cdot ln\left(\frac{sinh(C/T)}{sinh(C/T_{ref})} \right) - DE \left(\frac{tanh(E/T)}{T} - \right. \\ . & \left. \frac{tanh(E/T_{ref})}{T_{ref}} \right) + D \cdot ln\left(\frac{cosh(E/T)}{cosh(E/T_{ref})} \right) \\ . & - R \cdot log\left(\frac{P}{P_{ref}} \right) \end{aligned} \tag{30}$$

In Eq. (8) and (30) A, B, C, D and E are the heat capacity constants according to Aly and Lee (1981).

The departure function of entropy for the Peng-Robinson equation is shown in Eq. (31) [1]

$$s(p,T)^{PR\,dep} = R\left[T_r(Z-B) - 2.078m\left(\frac{1+m}{\sqrt{T_r}} - \kappa \right) \ln\left(\frac{Z+2.414B}{Z-0.414B} \right) \right] \tag{31}$$

A.4. General ORC model description

The model of an organic Rankine cycle (ORC) and was adapted and simplified from a study of Andreasen et al. (2014)[7].

The model describes a base case of an ORC that uses heat from a low temperature heat source, i.e. a hot water stream. It consists of a pump, a boiler (comprising a preheater, an evaporator and a superheater), a turbine and a condenser. The goal is to calculate temperature T_i, pressure P_i, enthalpy h_i and entropy s_i at every state i of the cycle subject to a given set of input parameters and variables. The model output is the net power produced by the cycle. The working fluid used in this base case is 1,1,1,3,3-Pentafluoropropane (R245fa).

The process overview is shown in figure 1 including the numbering of the streams and the corresponding variables. In order to calculate all the state variables an equation of state model is needed. The process is divided in at high pressure side (state 2 and 3) and a low pressure side (state 1 and 4). The equation of state used is the Peng-Robinson equation described in the previous section.

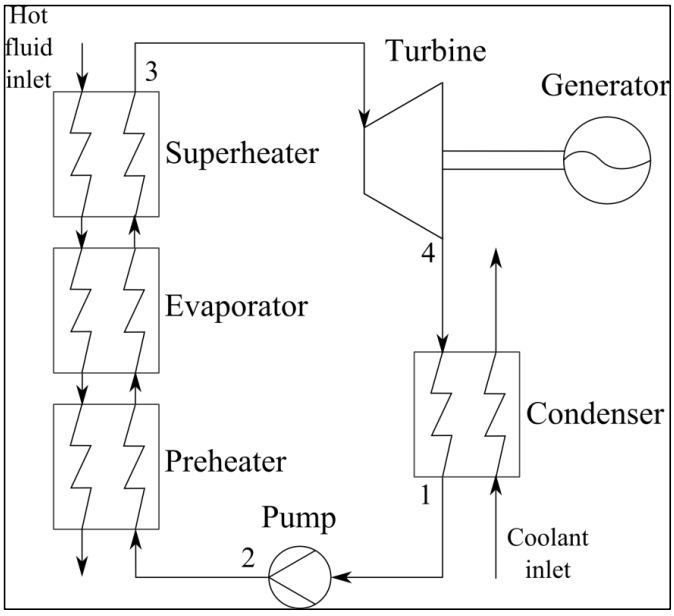


Figure 1: A sketch of the ORC process [7]

A.5. Component model

If two state variables are given, all other variables of the same state can be calculated using PR EOS. The basic equations for the process components are explained in the following. The complete model including all process equations, equations of states and process specifications is shown further below.

Pump model (state 1 -> 2)

The pump sucks the liquid from the condenser and increases its pressure, enthalpy and temperature. State one is assumed to be under saturation condition after the condenser. The pump model calculates first the isentropic state h_{2s} and subsequently the real state h_2 by applying the isentropic efficiency η_p.

$$P_1 = P_{sat}(T_1, T_c, P_c, \omega, A, B, C, D, E) \tag{32}$$

$$h_1 = h^l_{sat}(T_1, T_c, P_c, \omega, A, B, C, D, E) \tag{33}$$

$$s_1 = s^l_{sat}(P_1, T_1, T_c, P_c, \omega, A, B, C, D, E) \tag{34}$$

$$s_{2s} = s_1 \tag{35}$$

$$P_{2s} = P_2 \tag{36}$$

$$h_{2s} = h(P_{2s}, s_{2s}, T_c, P_c, \omega, A, B, C, D, E) \tag{37}$$

$$h_2 = h_1 + \frac{h_{2s} - h_1}{\eta_p} \tag{38}$$

$$T_2 = T(P_2, h_2, T_c, P_c, \omega, A, B, C, D, E) \tag{39}$$

Evaporator (state 2->3)

The boiler transfers heat from the low-temperature water heat source to the ORC system. In this model, the working fluid is assumed to evaporate completely. At the working fluid side (cold side) saturation conditions are assumed at the evaporator inlet (saturated liquid) and outlet (saturated vapor). The temperature of the heat source is obtained using the pinch point of the boiler $T_{boil\ pinch}$. As a first try, the location of the minimum pinch point temperature difference is assumed to be between the preheater and the evaporator. The procedure used for finding the heat source temperature at the location of the pinch point is listed below:

$$h_{hf,in} = h(P_{hf}, T_{hf,in})_{water} \tag{40}$$

$$P_2 = P_{pp,wf} \tag{41}$$

$$T_{wf,pp} = T_{sat}(P_{wf,pp}, T_c, P_c, \omega, A, B, C, D, E) \tag{42}$$

$$h_{wf,pp} = h^l_{sat}(P_{wf,pp}, T_c, P_c, \omega, A, B, C, D, E) \tag{43}$$

$$T_{hf,pp} = T_{wf,pp} + \Delta T_{boil\ pinch} \tag{44}$$

$$h_{hf,pp} = h(P_{hf}, T_{hf,pp})_{water} \tag{45}$$

Where subscripts *wf*, *hf* and *pp* denote working fluid, hot fluid and pinch point respectively.

The mass flow of the working fluid can be obtained by an energy balance over the evaporator and superheater.

$$\dot{m}_{wf} = \dot{m}_{hf} \cdot \frac{h_{hf,in} - h_{hf,pp}}{h_3 - h_{wf,pp}} \tag{46}$$

Furthermore the enthalpy of the outlet heat source stream is obtained by and overall energy balance.

$$h_{hf,out} = h_{hf,in} - \dot{m}_{wf} \cdot \frac{h_3 - h_2}{\dot{m}_{hf}} \tag{47}$$

$$T_{hf,out]} = T(P_{hf}, h_{hf,out})_{water} \tag{48}$$

The temperature difference between the working fluid and the hot fluid is then checked at the inlet to the preheater ($T_{hf,out} - T_2$). If the temperature difference is smaller than the pinch point temperature difference for the boiler, then the boiler calculation is repeated by assuming the pinch point to be located at the preheater inlet. If the temperature difference is larger than the boiler pinch point, then the procedure continues to the turbine model.

Turbine model (state 3->4)

The turbine expands the fluid and decreases subsequently its pressure and enthalpy while generating power. Similar to the pump model, first the isentropic enthalpy h_{4s} is obtained. Afterwards the real enthalpy value h_4 using the isentropic efficiency η_T is calculated.

$$h_3 = h(P_3, T_3, T_c, P_c, \omega, A, B, C, D, E) \tag{49}$$

$$s_3 = s(P_3, T_3, T_c, P_c, \omega, A, B, C, D, E) \tag{50}$$

$$s_{4s} = s_3 \tag{51}$$

$$P_{4s} = P_1 \tag{52}$$

$$h_{4s} = h(P_{4s}, s_{4s}, T_c, P_c, \omega, A, B, C, D, E) \tag{53}$$

$$h_4 = h_3 - (h_3 - h_{4s})\eta_T \tag{54}$$

$$T_4 = T(P_4, h_4, T_c, P_c, \omega, A, B, C, D, E) \tag{55}$$

Condenser (state 4->1)

In this simple pure component model, the states of the condenser are given by the specifications and equations of the other components.

Power output

The power output of the turbine and the power consumption of the pump are calculated from the energy balances over the two components. The net power output of the cycle is the difference between turbine and pump power.

$$\dot{W}_{TURB} = \dot{m}_{wf} \cdot (h_3 - h_4) \tag{56}$$

$$\dot{W}_{PUMP} = \dot{m}_{wf} \cdot (h_2 - h_1) \tag{57}$$

$$\dot{W}_{NET} = \dot{W}_{TURB} - \dot{W}_{PUMP} \tag{58}$$

Specifications

Usually the low temperature heat source is specified for a given problem. Further specification concern the pinch point temperatures of evaporators and the efficiencies of pump and turbine:

$\dot{m}_{hf}$, $T_{hf,in}$, η_p, η_T, $\Delta T_{boil\ pinch}$, T_1, P_{hf}

A.6. Comparison of sampling based optimization with particle swarm optimization

We suggested an alternative approach to identify the process variables (turbine input pressure P_t and temperature T_t). This was described in Step 3 in "2. Method and Tools" in the main manuscript. Here the results of the comparison between particle swarm optimization and the sampling based approach are shown.

A process-set up (corresponding to the same constraints and process parameters as described in the main manuscript) was solved for the working fluid R245fa. The process variables (turbine input pressure P_t and temperature T_t) were first determined using particle swarm optimization. Afterwards the identification of the optimal process variables was repeated using the sampling based approach. The process variables search space was sampled (using Latin hypercube sampling) and subsequently the model was evaluated for all of the (P_t, T_t)-samples. Finally, the process variables with the highest net power output were selected to be optimal. This procedure was repeated 20 times. The results are depicted as average relative deviation (ARD) between the solution from particle swarm optimization and the respective solution of the sampling based approach.

Table 1. Comparison of particle swarm optimization and sampling based optimization.

Average relative deviation ARD in %	*Value*
$\dot{W}_{NET}$	1.01
P_t	0.40
T_t	1.01

A.7. References for property model and ORC

[1] Liu W, Meinel D, Wieland C, Spliethoff H. Investigation of hydrofluoroolefins as potential working fluids in organic Rankine cycle for geothermal power generation. Energy 2014;67:106–16.
[2] Aly F, Lee L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. Fluid Phase Equilib 1981;6:169–79.
[3] Robinson DB, Peng D-Y, Chung SYK. The development of the Peng-Robinson equation and its application to a phase equilibrium system containing methanol. Fluid Phase Equilib 1986;24:25–41.
[4] Peng D-Y, Robinson DB. A New Two-Constant Equation of State. Ind Eng Chem Fundam 1976;15:59–64.
[5] Richard JE, Lira CT. Introductory chemical engineering thermodynamics. 2nd ed. Upper Saddle Ridge, NJ: Prentice Hall International Series in the Physical and Chemical Engineering Sciences; 1999.
[6] Smith JM, Van Ness HC, Abbott MM. Chemical engineering thermodynamics. 6th ed. McGraw-Hill Companies; 2001.
[7] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic Rankine cycles. Energy 2014;73:204–13.

A.8. References for REFPROP models

Fluid 1: Decafluorobutane (R610)

Huber, M.L. and Ely, J.F., A predictive extended corresponding states model for pure and mixed refrigerants including an equation of state for R134a, Int. J. Refrigeration, 1994;17:18-31.

Fluid 2: Octafluorocyclobutane (RC318)

Platzer, B., Polt, A., and Maurer, G., Thermophysical properties of refrigerants, Berlin, Springer-Verlag, 1990.

Fluid 3: 2-Chloro-1,1,1,2-Tetrafluoroethane (R124)

de Vries, B., Tillner-Roth, R., and Baehr, H.D., Thermodynamic Properties of HCFC 124, 19th International Congress of Refrigeration, The Hague, The Netherlands, International Institute of Refrigeration, IVa:582-589, 1995.

Fluid 4: Perfluoro-n-Pentane

Huber, M.L. and Ely, J.F., A predictive extended corresponding states model for pure and mixed refrigerants including an equation of state for R134a, Int. J. Refrigeration, 1994;17:18-31.

Fluid 5: Isobutane

Buecker, D. and Wagner, W., Reference Equations of State for the Thermodynamic Properties of Fluid Phase n-Butane and Isobutane, J. Phys. Chem. Ref. Data, 2006; 35:929-1019.

Fluid 6: N/A

Fluid 7: Neopentane

Lemmon, E.W. and Span, R., Short Fundamental Equations of State for 20 Industrial Fluids, J. Chem. Eng. Data, 2006; 51:785-850.

Fluid 8: n-Butane

Buecker, D. and Wagner, W., Reference Equations of State for the Thermodynamic Properties of Fluid Phase n-Butane and Isobutane, J. Phys. Chem. Ref. Data, 2006 35:929-1019.

Fluid 9: 1-Chloro-1,1-difluoroethane (R142b)

Lemmon, E.W. and Span, R., Short Fundamental Equations of State for 20 Industrial Fluids, J. Chem. Eng. Data, 2006M 51:785-850.

Fluid 10: Isobutene Lemmon, E.W. and Ihmels, E.C., Thermodynamic Properties of the Butenes. Part II. Short Fundamental Equations of State, Fluid Phase Equilibria, 2005; 228-229C:173-187.

Fluid 11: Trifluoroiodomethane

McLinden, M.O. and Lemmon, E.W. Thermodynamic Properties of R-227ea, R-365mfc, R-115, and R-13I1, J. Chem. Eng. Data, 2013.

Fluid 12: 1-Butene

Lemmon, E.W. and Ihmels, E.C., Thermodynamic Properties of the Butenes. Part II. Short Fundamental Equations of State, Fluid Phase Equilibria, 2005, 228-229C:173-187.

Fluid 13: 1,1,1,2,3,3-Hexafluoropropane (R236ea)

Rui, X., Pan, J., Wang, Y. An Equation of State for Thermodynamic Properties of 1,1,1,2,3,3-Hexafluoropropane (R236ea), Fluid Phase Equilib., 2013.

Fluid 14: 1,1,1,2,3,3,3-Heptafluoropropane (R227ea)

McLinden, M.O. and Lemmon, E.W. Thermodynamic Properties of R-227ea, R-365mfc, R-115, and

R-13I1, J. Chem. Eng. Data, 2013.

Fluid 15: Heptafluoropropylmethylether

Zhou, Y. and Lemmon, E.W. preliminary equation, 2012.

Fluid 16: 1,1,1,3,3-Pentafluoropropane (R245fa)

Lemmon, E.W. and Span, R., Short Fundamental Equations of State for 20 Industrial Fluids, J. Chem. Eng. Data, 2006; 51:785-850.

Fluid 17: trans-2-Butene

Lemmon, E.W. and Ihmels, E.C., Thermodynamic Properties of the Butenes. Part II. Short Fundamental Equations of State, Fluid Phase Equilibria, 2005; 228-229C:173-187.

Fluid 18: 1,2-Dichlorotetrafluoroethane (R114)

Platzer, B., Polt, A., and Maurer, G., Thermophysical properties of refrigerants," Berlin, Springer-Verlag, 1990.

Fluid 19: N/A

Fluid 20: 1,1,1,3,3,3-Hexafluoropropane (R236fa)

Pan, J., Rui, X., Zhao, X., Qiu, L. An equation of state for the thermodynamic properties of

1,1,1,3,3,3-hexafluoropropane (HFC-236fa), Fluid Phase Equilib., 2012; 321:10-16.

Paper G

Reverse engineering of working fluid selection for industrial heat pump based on Monte Carlo sampling and uncertainty analysis

Jérôme Frutiger[a], Benjamin Zühlsdorf[b], Brian Elmegaard[b], Jens Abildskov[a], Gürkan Sin[a]

a) Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
b) Department of Mechanical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

in preparation, to be submitted to Energy (ISSN: 0360-5442), 2017.

Abstract

This study presents a novel methodology for the identification of suitable pure component working fluids for heat pumps. Two challenges are addressed: the difficulties in finding a feasible (real fluid) solution when solving a product-process design problem, and the impact of the working fluid property uncertainties on the solution. A Monte Carlo sampling is applied to generate sets of different property parameter combinations (virtual fluids), which are subsequently evaluated in the heat pump process model. The distance between the property values of the virtual fluid and the uncertainty bound of the properties of real fluids (collected from a database) are calculated. The fluids that are closest to the top performing virtual fluids are further analyzed through evaluation in the cycle and subsequent uncertainty propagation of the respective input property uncertainties to the model output uncertainties. The methodology has been applied to an industrial heat pump system used for waste heat recovery from spray drying facilities in dairy industry. To remain focused on the validation of underlying concepts of the methodology, the study considered cyclic hydrocarbon working fluids. The compounds identified by the methodology had a low global warming potential, but a high flammability, where cyclopentane showed the best performance. The sampling based reverse engineering method identified top performing working fluids, but avoided solving computationally demanding molecular design problems and took into account the real fluid property uncertainties.

1. Introduction

The integration of computational chemical product and process design has become a widely used principle in computer-aided process engineering, in order to identify the most suitable process chemicals and simultaneously optimize the process conditions [1]. For example product-process design has been used in a variety of applications such as for the development of solvents [2] and surfactants in chemical separation [3], or for active ingredients for drug discovery in pharmaceutical engineering [4]. Most recently promising working fluids (both pure components and mixtures) for heat pumps and organic Rankine cycles (ORC) were identified using computer-aided product and process design techniques [5].

Computational chemical product design problems are based on a reverse engineering approach, where chemical compounds are identified based on optimal target properties required by the process. Molecules with optimal target properties can either be constructed molecules, as in computer-aided molecular design (CAMD), or identified in databases [6] as in classical reverse engineering approach. To this extent the chemical property values can either be predicted (e.g. using group contribution methods) [7] or obtained from databases containing experimental data [8][9][10].

We would like to pay attention to two major challenges with respect to property models and computational product design: 1) the difficulties in finding a feasible (real fluid) solution when solving a complex product-process design optimization problem, and 2) the property uncertainties, caused by the measurements [11] or by the property prediction models [12].

The combination of accurate property models and complex process equations usually leads to optimization problems with a large amount of non-linear and/or mixed-integer constraints and equations [13]. The solution of these computationally demanding problems requires advanced solvers [14] and is usually time-consuming. Furthermore, the globally optimal solution (i.e. a molecule defined by a combination of target properties) is not necessarily a feasible solution, since the property search space is not continuous [15]. And there might be multiple local

optima that may be advantageous in terms of practical feasibility compared to the actual global optimum. Hence, even though the complex optimization is solved successfully, tedious problem reformulations or post-screening may be necessary in practice. The non-continuous search space also makes classical contour mapping of the property search space [16] unsuitable as an analysis tool for the product-process design problem. Further, when a set of target properties is identified, the determination of the influence (or importance) of these properties on the process model output often relies on expert knowledge [17][18].

The problem of property uncertainties in the context of chemical product design has already been addressed by several authors, and the work of Ng et al. [19] gives an overview about the studies carried out in this domain. In particular, Maranas [20] used multivariate probability density functions for chemical properties incorporated into the problem formulation. Diwekar et al. [21] developed algorithms using uncertainty factors [21]. Fuzzy optimization has been used by Ng et al. [22][23] to take into account property uncertainties in connection with molecular design. One of the main challenges when taking into account property uncertainties is that the product design results in the formulation of a multistage problem, which is unusually computationally intensive to solve [19]. Hence, taking uncertainties into account usually increases the difficulties in finding a feasible solution to a reverse engineering problem.

In the field of application of reverse engineering approaches and molecular design, the identification of novel working fluids for thermodynamic cycles has been investigated by several authors. In particular new pure component and mixture working fluids for organic Rankine cycles and heat pumps from low-temperature heat sources [24] have been of particular interest due to the large amount of waste heat available in industry [25]. The choice of the working fluid has a significant influence on the thermodynamic and economic performance of a heat pump or an ORC. The thermodynamic properties of the working fluid strongly influence the cycle design and thereby the thermodynamic performance. Hence, the identification of a suitable working fluid is crucial for the (overall economic) performance/feasibility of the heat pump. Nevertheless, the choice of the working fluid is restricted by legislation limiting the use of environmentally harmful fluids.

The reviews of Bao et al. [26] on fluid selection and Linke et al. [5] on molecular design, give an inclusive summary of studies considering working fluids for ORCs and heat pumps. Concerning algorithms for the identification or design of working fluids, the following studies should be considered in particular: Brown et al. [27][28] investigated theoretically ideal working fluids through the variation of fluid properties. Stijepovic et al. [29][30] analyzed the relation between working fluid properties and economic performance criteria. In similar way a number of studies investigated the influence of a certain fluid property (e.g. the critical temperature) on the performance of a working fluid [31][32][33][34][15] in a classical one-factor-at-the-time approach of sensitivity analysis. Papadopoulos et al. [35][36] presented a method to identify and quantify the influence of a number of model parameters in the design of mixtures. However, apart from the this work, the influence of the working fluid target properties has not yet been identified by a global sensitivity analysis technique [37] and combined with the identification of new working fluids.

Computational molecular design techniques were used by Papadopoulos et al. [38][35][39] using multi-objective optimization in order to find optimal molecular structures of pure components as well as the composition of mixtures for ORC processes. Furthermore, Palma-Flores et al. [40] and Cignitti et al. [41] defined the product-process design problem for the working fluid selection as mixed integer non-linear programming (MINLP) optimization problem. In a similar way Molina-Thierry et al. [42] performed a constraint simultaneous optimization of fluid mixtures (generated from a pre-specified set of pure fluids) and the

process operating conditions. Lampe et al. [43][44] performed fluid-searching using a reverse engineering approach for ORC processes. Roskosch et al. [16] applied a reverse engineering approach for a heat pump application. Roskosch et al. tried to thoroughly analyze the fluid search space by contour mapping, in order to get insight into the sensitivities of the fluid properties in the product-process design model. They showed that the critical pressure of a working fluid had only a minor influence on the heat pump performance.

All the above-mentioned studies on working fluid design tried to solve a complex and computationally demanding integrated optimization problem and only a few studies [35][36] considered explicitly the overall influence of the uncertainties of fluid properties on the model output.

In this work, we address the discussed challenges in product-process design on the difficulties in finding a feasible (real fluid) solution and on taking into account property uncertainties when solving the mathematically complex problem formulation. We present a novel methodology for reverse engineering based on Monte Carlo sampling and uncertainty analysis and show its application to the identification of novel working fluids for an industrial heat pump system.

A Monte Carlo based method has recently been presented by Frutiger et al. [45][46] as a methodology to propagate and quantify the impact of property parameter uncertainties to the process model output as well as to assess equations of state based on uncertainties. Monte Carlo sampling and model evaluation was used to propagate the uncertainty of the fluid-specific property parameters to the thermodynamic cycle model output. This methodology provided distributions of the model output resulting from fluid property uncertainties. The study also showed that taking into account property uncertainties in a product-process design problem allows interpreting the ranking of the product solutions in different ways: An optimistic ranking considers the higher bound of the model output distribution for the different fluids and therefore ranks the fluids according to their highest possible potential. A more conservative approach, on the other hand, ranks the fluids according to the lower bound or the mean value of the output uncertainty distribution.

The methodology described in this work addresses in particular the following aspects that have not attracted sufficient attention in the literature:

- Usage of simple and computationally efficient stochastic methods such as Monte Carlo sampling to identify optimal sets of working fluid target properties in a reverse engineering problem for product-process design
- Global sensitivity analysis of the property search space to investigate the overall influence of a particular target property on the process model output
- Identification of real working fluids based on the optimal target properties and taking into account the real fluid property uncertainties
- Focus on the interpretation of the product-process model results including its uncertainties and avoiding the solution of a time-consuming and complex optimization problem

We apply the methodology to an industrial case study: a heat pump system for waste heat recovery from a spray drying facility for milk powder production [47]. In food industries spray drying processes are highly energy intensive and are responsible for a large fraction of waste heat, which is often rejected into the environment [48]. A heat pump can be used to recover the waste heat from the exhaust air of a spray dryer and re-utilize it in the pre-heating of the air entering the dryer. In this way the power consumption of the industrial facilities can be

significantly reduced [47]. The goal of the case study is to identify suitable novel working fluid candidates, which allow a high performance of the heat pump.

The methodology consists of the formulation of the product-process model for the working fluid identification of a heat pump. The working fluid target property search space is specified and Monte Carlo sampling is applied to generate sets of different property parameter combinations (virtual fluids), which are subsequently evaluated in the heat pump process model. Furthermore, a derivative-based global sensitivity analysis is performed to analyze the influence of the target properties on the model out. Afterwards real fluids that lie closest to the top-performing virtual fluids are identified taking into account the fluid property uncertainties and the sensitivity of the respective properties on the heat pump model output.

The paper is organized as follows: (i) the overall methodology is outlined; (ii) the heat pump model including the used property models is presented (iii) the Monte Carlo procedure for sampling and exploration of the target property search space is explained (iv) the global sensitivity analysis method is described (v) the algorithm for the identification of suitable working fluid candidates is presented (vi) the results for the working fluids are compared considering the heat pump performance together with safety-related and environmental properties.

2. Method and Tools

An overview of the procedure divided in different steps is shown in Table 1.

Table 1. Overview of the procedure.

Step 1	Formulation of model for heat pump and thermodynamic property estimation
Step 2	Specification of working fluid property descriptors and search space for reverse engineering
Step 3	Generation and evaluation of virtual fluids: Monte Carlo based sampling of property search space and evaluation in process model
Step 4	Global sensitivity analysis of working fluid property descriptors and identification of property weights
Step 5	Calculation of distance function between properties of real and virtual fluids and ranking of real fluids
Step 6	Evaluation of identified high-ranked real working fluids including uncertainty analysis

2.1. Step 1: Formulation of model for heat pump and thermodynamic property estimation

In this study optimal pure component working fluid candidates for an industrial heat pump system are to be identified. The reverse engineering system can be considered as an combined product-process design problem [14], where the most suitable chemical products (i.e. working fluids) are identified simultaneously with the optimal process conditions (i.e. the heat pump

parameters). The heat pump of this study is a waste heat recovery (WHR) system used in a spray drying facility for milk powder production in a reference dairy factory located in Denmark [25]. The heat pump system is used to recover heat from exhaust gas from the spray dryer (heat source) to preheat the air before the spray dryer (heat sink), utilizing secondary cycles with pressurized water. The heat pump is outlined in Figure 1. It consisted of four main components: a compressor, a condenser, a throttling valve and an evaporator. Heat is provided by the water source to the working fluid through the evaporator (state 6 to 7 in Figure 1) and the super-heater (state 7 to 1). Afterwards the working fluid is pressurized by the compressor (state 1 to 2) and the heat is rejected to the sink water via the desuper-heater, the condenser and the sub-cooler (states 2 to 5). A throttling valve expanded the working fluid before re-entering the evaporator (state 5 to 6). The heat pump process model and constraints are based on the work of Zühlsdorf et al. [47].

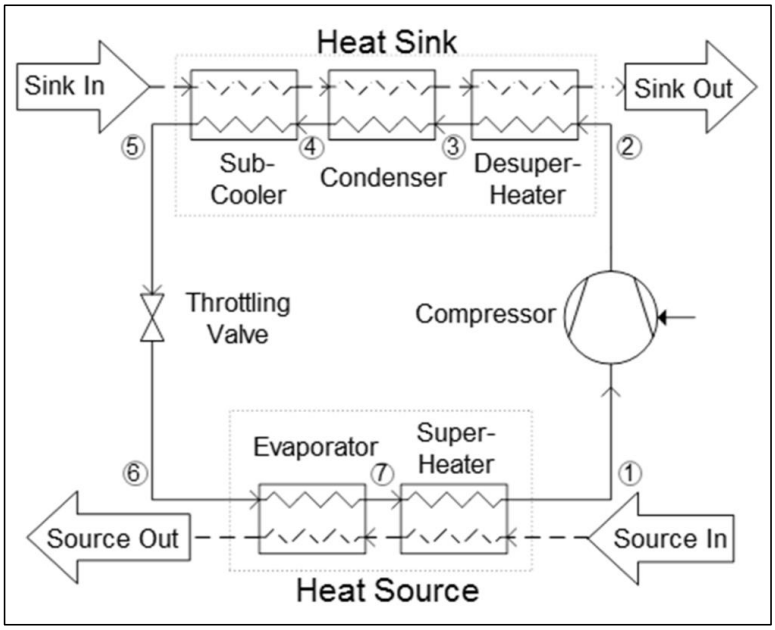


Figure 1. An outline of the industrial heat pump for waste heat recovery [47].

The boundary conditions and additional assumed parameters for the heat pump simulations are listed in Table 1. The heat source is water at a temperature of 65 °C and a mass flow of 14.18 kg/s, i.e. the waste heat stream of the dairy factory. The heat sink is a water stream of 10.61 kg/s, being heated up from 75 °C to 125°C. There are no limitations imposed on the source outlet temperature, since it is defined by energy balances. Further, the heat capacities of the water flows are assumed to be constant.

Table 1. Boundary Conditions for the heat pump system [47].

Process parameter	*Value*
Source water inlet temperature	65 °C
Source water mass flow	14.18 kg/s

Source water heat capacity	4.18 kJ/kgK
Sink water inlet temperature	75 °C
Sink water outlet temperature	125°C
Sink water mass flow	10.61 kg/s
Sink water heat capacity	4.21 kJ/kgK
Compressor efficiency	0.8
Compressor motor efficiency	0.95

In the numerical modeling the following assumptions are used: homogeneous flow in terms of thermodynamic properties, no heat loss from the system, steady state condition and no pressure losses in piping or heat exchangers [45].

In the cycle simulation the thermodynamic properties (i.e. specific enthalpies and entropies) of the working fluid need to be calculated at each state. An equation of state (EoS) is required consisting of an ideal contribution (i.e. the ideal gas enthalpy and entropy) and a departure function accounting for the difference between ideal and real behaviour. Peng-Robinson Equation of State (PR-EoS) [49] is selected to determine the departure functions of the thermodynamic properties. PR-EOS has commonly been used for design or screening of working fluids [45][27][50]. The main advantage of PR-EOS compared to forms of the high accuracy Helmholtz EoS [51][52], is its relatively small number of required fundamental parameters as a 3-parametric cubic equation of state.

The PR-EoS is given by,

$$P = \frac{RT}{V_m - b} - \frac{a\alpha}{V_m^2 + 2bV_m - b^2} \tag{1}$$

In Eq. (1) R is the universal gas constant, T is the absolute temperature, P is the pressure, V_m is the molar volume. The other parameters are defined as follows

$$a = \frac{0.457235 R^2 T_c^2}{P_c} \tag{2}$$

$$b = \frac{0.077796\, R\, T_c}{P_c} \tag{3}$$

$$\alpha = \left(1 + 0.37464 + 1.54226\,\omega - 0.26992\,\omega^2 \left(1 - \sqrt{\frac{T}{T_c}}\right)\right)^2 \tag{4}$$

In Eq. (2) to (4) T_c is the critical temperature of the working fluid, P_c is the critical pressure, ω is the acentric factor. In order to solve PR-EOS only three fluid-specific parameters are required: T_c, P_c and ω. The computational implementation of the PR EoS is based on the work of Frutiger et al. [45], who provided the detailed property model description in their work.

The ideal-gas enthalpy and entropy terms of the EoS are obtained from the integration of a temperature-dependent ideal-gas heat capacity function, $c_p(T)$. As suggested by Roskosch et al.

[16], we use a linear temperature dependence of the isobaric ideal gas heat capacities (expressed by two parameters *A* and *B* only) in order to lower the amount of fluid specific parameters for the reverse engineering problem. The heat capacity correlation is shown in Eq. (5).

$$c_p(T) = A + B \cdot T \tag{5}$$

A linear dependence of a small temperature range as in the given heat pump, shows to be sufficiently accurate [16] in the temperature range of the cycle.

The outputs of the heat pump model are the coefficient of performance *COP* (i.e. the ratio between the supplied heat and the compressor power input to the system), the working fluid mass flow $\dot{m}_{\mathrm{wf}}$, and state variables such as pressures P_i, temperatures T_i, entropies s_i, and enthalpies h_i, (see Figure 1).

The *COP* is calculated from Eq. (6)

$$COP = \frac{\dot{Q}_{\mathrm{sink}}}{\dot{W}} \tag{6}$$

where $\dot{Q}_{sink}$ is the heat provided to the water sink and $\dot{W}$ is the compressor power. $\dot{Q}_{sink}$ and $\dot{W}$ are expressed as:

$$\dot{Q}_{\mathrm{sink}} = \dot{Q}_{\mathrm{source}} + \dot{W} \tag{7}$$

$$\dot{W} = \dot{m}_{\mathrm{wf}} \cdot (h_2 - h_1) \tag{8}$$

h_1 and h_2 are the enthalpies at state 1 and 2. The source heat $\dot{Q}_{source}$ can be written as in Eq. (9). The mass flow $\dot{m}_{\mathrm{wf}}$ is calculated according to Eq. (10).

$$\dot{Q}_{\mathrm{source}} = \dot{m}_{\mathrm{source}} \cdot c_{\mathrm{p\ source}} \cdot (T_{\mathrm{source\ in}} - T_{\mathrm{source\ out}}) \tag{9}$$

$$\dot{m}_{\mathrm{wf}} = \frac{\dot{Q}_{\mathrm{source}}}{h_1 - h_6} \tag{10}$$

In Eq. (9) $\dot{m}_{\mathrm{source}}$ expresses the source water mass flow, $c_{\mathrm{p\ source}}$ is its corresponding heat capacity. $T_{\mathrm{source\ in}}$ and $T_{\mathrm{source\ out}}$ are the source input and output temperatures. In Eq. (10) h_1 and h_2 are the enthalpies at state 1 and 6.

A degree of freedom analysis shows that the cycle can be solved by fixing or optimizing two process variables, which are chosen to be the condensation and the evaporation pressure. The optimal process conditions for each set of property descriptors (i.e. virtual fluids) are identified by Newton–Raphson method [53] with the *COP* as the objective function. The optimization problem in this study is much simpler compared to other studies, that solve e.g. MINLP problems [40].

2.2. Step 2: Specification of working fluid property descriptors and search space for reverse engineering

The fluid-specific property descriptors [54] for the reverse engineering problem need to be specified. For the given case study the property input parameters to the ideal-gas contribution and departure function of the EoS are chosen. These are:

- heat capacity correlation constants: A and B
- critical temperature T_c
- critical pressure P_c
- acentric factor ω
- molecular weight MW

For each of the six property descriptors a value range (lower and upper bound) needs to be specified. The six ranges together form a six-dimensional property search space. In the current case study the ranges are selected by analyzing property data from the well-established DIPPR 801 AIChE database [10]. The database contains data for T_c, P_c and ω for 190 refrigerants. Using percentile calculations of the respective data sets for T_c, P_c, ω and MW ranges of the values are obtained. The lower bounds corresponds to the 2.5% percentile of the DIPPR data set, whereas the upper bound corresponds to the 97.5% percentile. This means the ranges cover the data values in which 95% of the compounds lie in. This procedure is applied, in order to exclude very high or low values.

In the DIPPR database the temperature-dependent ideal gas heat capacity is described by the five-parametric Aly-Lee the correlation [55]. In order to obtain the parameters A and B of the simplified two-parameter correlation of Eq. (5), the temperature-dependent ideal gas heat capacities of the DIPPR database compounds are fitted by non-linear regression [56] to Eq. (5) in the temperature range of the heat pump, i.e. 40 °C to 150 °C. Afterwards, the same percentile calculations (as for T_c, P_c and ω) are applied to define the range of the parameters A and B.

Table summarizes the ranges of the fluid specific descriptors (search space).

Table 2. Search space for the fluid specific descriptors.

Property descriptor	*search space (range)*
heat capacity correlation constants A	0-70364 J/(kmolK)
heat capacity correlation constants B	10-500 J/(kmolK^2)
critical temperature T_c	365-620 K
critical pressure P_c	2300-12070 kPa
molecular weight MW	20-255 g/mol
acentric factor ω	0.05-0.9

The aim of the reverse engineering approach for fluid design is to identify best combinations of these property descriptors within the respective search space that provide an optimal heat pump model output, i.e. a high COP value.

2.3. Step 3: Generation and evaluation of virtual fluids: Monte Carlo based sampling of property search space and evaluation in process model

Monte Carlo based sampling within the property search space is used to generate different sets of fluid-specific descriptors. These sets essentially can be considered as virtual fluids representing the search space.

In this study a Quasi-Monte Carlo method (QMC) [57] is utilized for low-discrepancy sampling. Monte Carlo methods based on random number sampling is exposed to clustering of sampling points, which leads to a non-uniform reflection of the search space. Low-discrepancy sampling overcomes this issue using uniformly distributed sequences (i.e. Halton sequences [58]) and increases the uniform sampling of the space significantly.

From the fluid descriptor search space a total of 400 samples are generated.

Figure 2 provides an illustration of the sampling results.

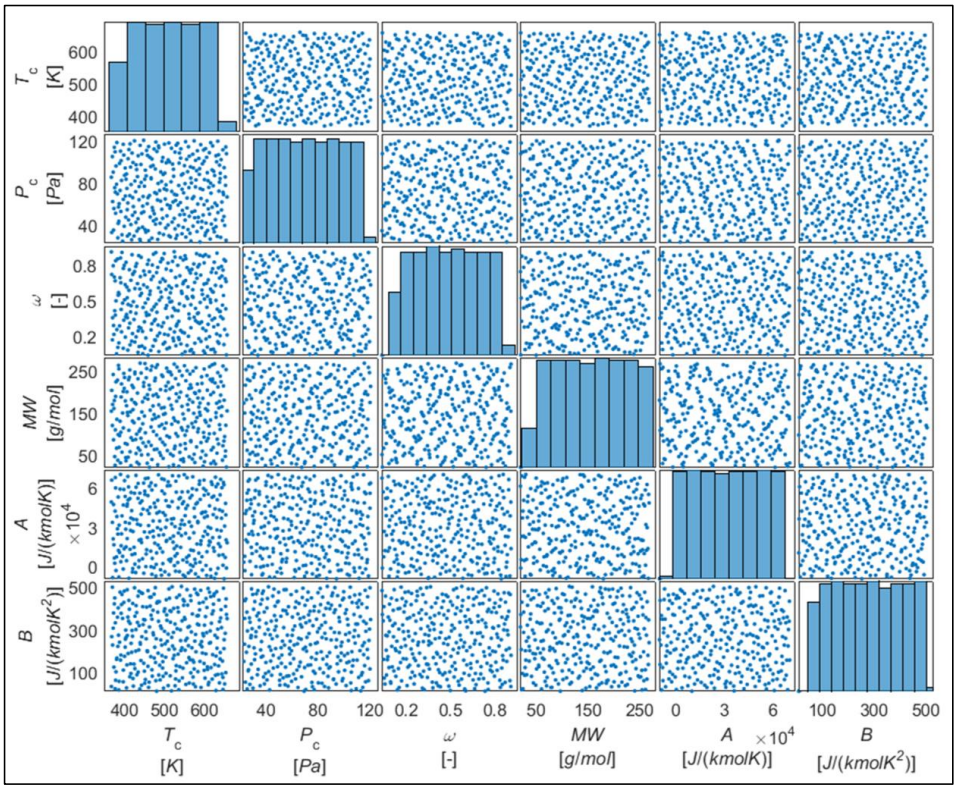


Figure 2. Illustration of samples generated by low-discrepancy sampling using Halton sequences: The sample matrix represents the property search space. The diagonal elements of the matrix represent the uniform distribution of the sampling.

One model simulation is performed for each of the 400 virtual fluids (molecular descriptor samples). For each of the virtual fluids the Newton-Raphson algorithm is used to obtain the optimal condensation and evaporation pressure. Virtual fluids that consist of property parameters infeasible with respect to the property model evaluations are screened out. Subsequently the virtual fluids are ranked according to their *COP*.

2.4. Step 4: Global sensitivity analysis of working fluid property descriptors and identification of property weights

A global sensitivity analysis of the property descriptors (A, B, T_c, P_c, ω) with respect to the model output (i.e. *COP*) allows analyzing the impact of the respective parameters in the heat

pump model [15]. This gives a deeper understanding of the effects that a change in a property parameter value can have on the heat pump model output. The property values of promising real working fluids should be as close as possible to the corresponding values of the best virtual fluids. As it will be shown in Step 7, a distance function can be calculated to identify the closeness of a virtual and a real fluid. However, depending on the influence of a certain property descriptor on the overall model output, the change in property value from the virtual to the close by real fluid may have a strong or weak effect on the model output. This means, property parameters can have a high or low sensitivity with respect to the heat pump model output, i.e. the *COP*. If they have a high sensitivity, the respective property descriptor should have a higher weight (importance factor), when calculating the distance function, than the ones with a low sensitivity.

We use derivative-based global sensitivity measures [37] to investigate the overall influence of the property descriptors in the search space. The detailed procedure for derivative-based global sensitivity analysis can be found in the work of Kucherenko et al. [37].

For each virtual fluid in the search space the local sensitivity of the respective property parameter with respect to the *COP* is calculated according to Eq. (11). The sensitivity analysis is here shown for the example of ω.

$$s_{\omega i} = \frac{\partial COP}{\partial \omega_i} \approx \frac{COP_{\mathrm{f}i}(\omega_{\mathrm{f}i}) - COP_{0i}(\omega_{0i})}{\omega_{\mathrm{f}i} - \omega_{0i}} \tag{11}$$

where $s_{\omega i}$ is the local sensitivity measure (forward derivative) and ω_i is the respective fluid property sample value ω of virtual fluid *i*. In total analogy the local sensitivity measure can be calculated for *A*, *B*, T_c and P_c. The forward derivative value is obtained by perturbing the corresponding property descriptor value by ξ=0.01:

$$\omega_{\mathrm{f}i} = \omega_{0i} \cdot (1 + \xi) \tag{12}$$

The local sensitivity measure is calculated at every sample (virtual fluid) data point in the search space. Hence, for every property descriptor local sensitivity measures are obtained over the whole search space.

The different sensitivity distributions can be summarized in the global sensitivity measure (expressed for the example of ω as S_ω^{tot}) using numerical integration of the respective distribution:

$$S_\omega^{\mathrm{tot}} = \frac{\int_i s_\omega \mathrm{d}x}{\pi^2 D_{COP_\omega}} \tag{13}$$

where s_ω is the distribution of the local sensitivity measure for the respective property ω (obtained from Eq. (12)), π is 3.14 and D_{COP_ω} is the variance of the *COP* values obtained from evaluating the different samples and serves as a normalization factor. The statistical explanation and derivation of Eq. (13) can be found in the work of Sobol et al. [59]

The global sensitivity measures of the corresponding property descriptors are ranked and a ranking value is assigned accounting for its importance. If two global sensitivity measures of two property descriptors lie in the same order of magnitude, the two property descriptors obtain the same ranking value. Afterwards, a normalized weight factor *w* is calculated by normalizing the ranking values (see Table 3). These weight factors are used in Step 7 for the calculation of the distance function.

Table 3. Global sensitivity measure S^{tot} and normalized weight factor w for the

Property descriptor	*Global sensitivity measure* S^{tot}	Importance Ranking	Normalized weight factor $w = rank\{i\} / \sum rank\{i\}$
acentric factor ω	132	5	0.28
Molecular weight MW	$1.58 \cdot 10^{-3}$	4	0.22
heat capacity correlation constants B	$2.12 \cdot 10^{-4}$	3	0.17
critical temperature T_c	$1.26 \cdot 10^{-4}$	3	0.17
heat capacity correlation constants A	$1.43 \cdot 10^{-7}$	2	0.11
critical pressure P_c	$4.34 \cdot 10^{-14}$	1	0.05

2.5. Step 5: Calculation of distance function between properties of real and virtual fluids and ranking of real fluids

A database of real chemical compounds needs to be selected. To compare the best performing virtual fluids results to real fluids. In this study we select the thermophysical property database ”ThermoData Engine” (TDE) [8][9] of the National Institute of Standards and Technology (NIST), which contains property data of over 20000 chemicals. A major advantage of TDE is that every data point is reported along with a corresponding uncertainty value (i.e. the 95% confidence interval).

Alternatively, a computer-aided molecular design (CAMD) algorithm [50] could be used and the best performing virtual fluids could be considered as target properties. In this way molecules could be generated that are close to the desired target properties.

In the current case study we focus on purely hydrocarbon based working fluids. While the methodology itself is generic and can be applied to screen larger classes of chemicals database, the latter focus is made intentional so as to remain focused on demonstrating proof of concept of the new methodology. In the context of the phase-out of chlorinated and fluorinated compounds for thermodynamic cycles in Europe [60], hydrocarbon based (natural) refrigerants show promising performances have no ozone depletion potential, and possess much lower global warming potential [61]. Furthermore, the TDE database contains a large number of cyclic hydrocarbons. Fluids of such type (e.g. cyclopentane) are often not considered when performing fluid design with “classical” computer-aided molecular design optimization algorithms due to the high number combinatorial possibilities and the difficulties of estimating property data for such compounds [40]. Hence, cyclic hydrocarbons have not been investigated thoroughly as working fluids. In order to illustrate the application of the new methodology, novel pure component cyclic hydrocarbons should be identified.

If experimental data is not available for P_c, T_c or ω for a particular compound, group contribution methods developed with the methodologies of Frutiger et al. [7][62] are used to estimate the properties. Every predicted data point is reported along with its corresponding uncertainty (95% confidence).

In total 2126 real hydrocarbon based fluids with property data including uncertainty range are used as real fluids for calculating the distance to the best performing virtual fluids (from Step 3) in the property search.

The distances in the search space between the property value of a top performing virtual fluid and a real fluid is calculated including the property uncertainty range of the respective real fluid. We show the calculation of the distances for the example of the acentric factor ω. The acentric factor of a real fluid *y* is considered as

$$\omega_y^{\text{low}} < \omega_y < \omega_y^{\text{up}} \tag{14}$$

where ω_y is the database value of the acentric factor ω for compound *y*, ω_y^{low} is the lower bound of the uncertainty range for the database value and ω_y^{up} is the upper bound respectively, as it is reported by TDE with 95% confidence.

We define the distance between a top performing virtual fluid *x* and a real fluid *y* in the property search space of the acentric factor ω as distance d_{xy}^{ω}. We use the following algorithm to assign and calculate the values of d_{xy}^{ω}:

a) If the acentric factor ω_x of the top performing virtual fluid *x* lies within the upper and lower bound of the uncertainty range of the acentric factor ω_y of real fluid *y*, the distance function d_{xy}^{ω} is assigned a zero value:

$$\text{if} \quad \omega_y^{\text{low}} < \omega_x < \omega_y^{\text{up}} \quad \text{then} \quad d_{xy}^{\omega} = 0 \tag{15}$$

b) If the virtual property ω_x is below the lower bound ω_y^{low} of the real property, then the normalized distance between ω_x and ω_y^{low} is calculated as follows

$$\text{if} \quad \omega_x < \omega_y^{\text{low}} \quad \text{then} \quad d_{xy}^{\omega} = \frac{|\omega_y^{\text{low}} - \omega_x|}{\omega_y^{\text{low}}} \tag{16}$$

where $|\omega_y^{\text{low}} - \omega_x|$ corresponds to the absolute value norm.

c) If the virtual property ω_x is above the upper bound ω_y^{low} of the real fluid, then the distance is obtained by

$$\text{if} \quad \omega_y^{\text{up}} < \omega_x \quad \text{then} \quad d_{xy}^{\omega} = \frac{|\omega_x - \omega_y^{\text{up}}|}{\omega_y^{\text{low}}} \tag{17}$$

The algorithm is repeated analogously for T_c and P_c to obtain d_{xy}^{Tc} and d_{xy}^{Pc}. For the molecular weight, there is no uncertainty range, hence to calculate d_{xy}^{MW} the distance to the actual real fluid value is taken.

The ideal gas heat capacity distance values between the virtual and the real fluids are calculated at 10 temperature points between 40 °*C* to 150 °*C* with the described algorithm. For the virtual fluids the simplified correlation $c_p(T) = A + B \cdot T$ is used. For the real fluids the TDE database tables for the ideal gas heat capacity are applied or, if not available, the heat capacity is predicted by the method of Joback and Reid [63]. The average distance over the temperature range is calculated for the ideal gas heat capacity giving one distance value d_{xy}^{cp}.

Figure 3 illustrates the principle for the calculation of the distance function for the two-dimensional search space of ω and T_c.

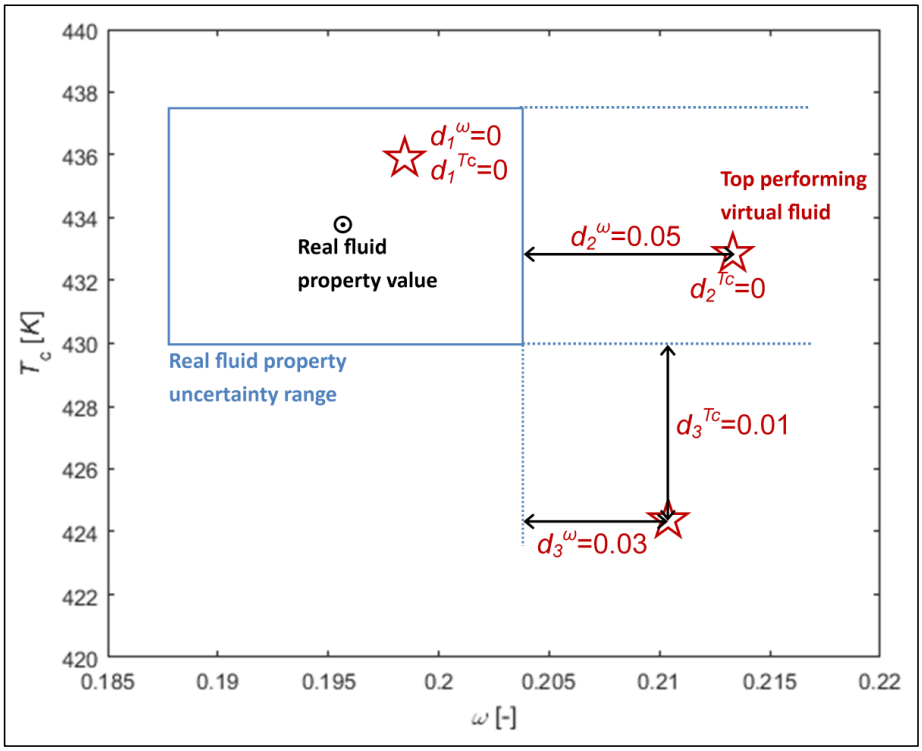


Figure 3. Illustration of the algorithm to calculate the distance function in the ω-T_c sub-search space: If the virtual fluid (stars) value lies within the uncertainty bounds of the respective real fluids (circle), the distance d_{xy} is zero. Otherwise the distance is calculated between the uncertainty bounds and the virtual fluid.

In Figure 3 the property values of a real compound (marker: circle) is shown with its corresponding uncertainty bounds together with three virtual top performing fluids (star). In the two-dimensional sub-search space the uncertainty bounds form a square around the corresponding property value. The virtual fluid 1 fell inside the uncertainty range. Hence its distance function values are zero for both ω and T_c. The T_c property value of virtual fluid 2 falls inside the uncertainty range, but its ω property value is higher than the upper bound of the real fluids. Therefore, the distance function needed to be calculated ω for between the upper bound and the value of virtual fluid. The property values of virtual fluid 3 lies completely outside the uncertainty ranges. This meant that the distance functions needs to be calculated for both T_c and ω.

The total distance function d_{xy}^{tot} for the distance of one virtual fluid x to one real fluid y is calcualted by summing up all the property distances multiplied by its corresponding weight factor w (obtained from Step 4).

$$d_{xy}^{\text{tot}} = d_{xy}^{\omega} \cdot w^{\omega} + d_{xy}^{Tc} \cdot w^{Tc} + d_{xy}^{Pc} \cdot w^{Pc} + d_{xy}^{MW} \cdot w^{MW} + d_{xy}^{\text{cp}} \cdot w^{\text{cp}} \tag{18}$$

In general form on can write the distance function for properties i from 1 to N as

$$d_{xy}^{\text{tot}} = \sum_{i}^{N} d_{xy}^{i} \cdot w^{i} \tag{19}$$

The acentric factor ω has the highest sensitivity and is assigned the largest normalized weight factor. Therefore, the acentric factor distance Eq. (18) d_{xy}^{ω} is penalized the strongest, whereas the critical pressure P_c has the smallest weight and is penalized the weakest. In other words, it is very important in the given heat pump system that the acentric factor of a real fluid is close to the acentric factor of a virtual fluid. However, it is less crucial that the critical pressures of the virtual and the real fluids are close to each other.

The distance function d_{xy}^{tot} is calculated between all the database compounds and the top 10% best performing virtual fluids.

The real working fluids are ranked according to the lowest total distance function value d_{xz}^{tot}.

2.6. Step 6: Evaluation of identified high-ranked real working fluids including uncertainty analysis

The high ranked real working fluids are evaluated in the heat pump model and a Monte Carlo based uncertainty analysis with respect to the property value uncertainties is performed for each of the compounds. To this extent the methodology of Frutiger et al. [45] for the propagation of the fluid property uncertainty to the heat pump model output is applied. The methodology uses Monte Carlo sampling of property values and subsequent evaluation in the cycle. However, unlike in Step 3 of the current study, the sampling does not take place over the whole search space, on the contrary, the samples are only generated within the uncertainty range of the respective property uncertainty. The methodology is briefly described here, details can be found in the work of Frutiger et al. [45]:

a) The sampling range is specified by the uncertainty (i.e. 95%-confidence interval) range of each real fluid property parameter (e.g. ω, T_c, P_c, c_p). A Monte Carlo sampling method, i.e. the Latin Hypercube sampling method [64], is utilized for probabilistic sampling of the properties within the uncertainty range. The probability of uncertainty is assumed to follow normal distribution. Correlations between PR EoS parameters (ω, T_c, P_c) are taken into account through the rank-based method for correlation control of Iman and Conover [65].

b) One model simulation of the heat pump system is performed for each of the property parameters samples. The Monte Carlo results provide a cumulative distribution function for the model output (i.e. *COP* value) of each of the real fluids. The distributions can be analysed using mean and percentile calculations. The 95%-confidence interval of the *COP* output with respect to the corresponding property parameter values can be obtained for every real fluid. This assesses the compounds not solely based on virtual fluid *COP*, or the actual fluid *COP*, but also including the uncertainty of the property data.

For comparison and validation, the results of some real fluids are compared to the *COP* obtained when using REFPROP database 9.0 [51].

The analysis of the output uncertainty of the *COP* for each real fluid allows to rank the promising fluid candidates according to the *COP* mean value, the lower bound (conservative approach) or the upper bound (optimistic approach).

Beside the heat pump model output value (*COP*) also environmental (global warming potential *GWP*) and safety-related properties (Lower flammability limit *LFL*) are calculated for the

respective real fluids. Therefore, the group contribution methods of Hukkerikar et al. [12] and Frutiger et al. [66] are used.

3. Results and Discussion

The results are shown in two parts. The results of the reverse engineering methodology (i.e. the identified compounds including their corresponding uncertainty) are presented and discussed. Afterwards the uncertainty analysis is further visualized for two compounds.

3.1. Ranking of working fluids including uncertainty

Table 4 shows the *COP* mean value of the distribution for the best performing compounds for the considered heat pump cycle. It includes the uncertainty with respect to the property input which was propagated through the cycle. Additionally the *COP* of the closest top performing virtual sample fluid is shown. The *COP* mean value and the corresponding uncertainty ranges (95%-confidence interval) are also represented in Figure 4. In order to compare the identified cyclic hydrocarbons, the *COP* has also been calculated for 3 commonly used refrigerants that would be suitable for the given process: R-152, R-143 and R-245fa. The ranking includes the lower flammability limit obtained from the prediction method of Frutiger et al. [66]. The global warming potential (GWP), calculated by the method of Hukkerikar et al. [67], is very small for all of the considered cyclic compounds (GWP<0.1) compared to commonly used fluoro-hydrocarbon refrigerants, as e.g. R-152 (GWP=53), R-143 (GWP=353), R-245fa (GWP= 1030) [68]. According to the recently published F-gas regulation of the European Union working fluids with a GWP of higher than 150 should be phased out [60]. Furthermore, the ozone depletion potential (ODP) of all of this cyclic hydrocarbon working fluids is zero [67].

Table 4. Best performing compounds ranked by *COP* mean value including uncertainty.

Rank	Working fluid name	Mean *COP*	*COP* 95%-conf. int.		*COP* of closest optimal virtual fluid	Sample number	LFL [Vol-%] [66] (with ASHRAE [69] Safety group)
1	Cyclopentane	3.06	3.00	3.11	3.17	3	1.41±0.90 (A3)
2	Cyclobutane	3.04	2.95	3.12	3.17	3	1.41±0.59 (A3)
3	cis-1,2-Dimethylcyclo-propane	3.04	2.87	3.11	3.17	3	1.40±0.58 (A3)
4	Methylcyclo-butane	3.03	2.86	3.11	3.17	3	1.78±0.90 (A3)
5	trans-1,2-Dimethylcyclo-propane	3.00	2.84	3.09	3.17	3	1.31±0.57 (A3)
6	Methylcyclo-propane	2.97	2.88	3.05	3.17	3	1.78±0.50 (A3)
7	1,1-Dimethyl-cyclopropane	2.86	2.63	3.08	3.19	1	0.69±0.59 (A3)
8	Butylcyclobutane	2.73	2.38	3.03	3.16	4	0.76±0.90 (A3)
9	1,1-cis-3,4-Tetramethylcyclo-pentane	2.49	2.25	3.01	3.18	2	0.51±0.60 (A3)
10	1,1,2-Trimethyl-2-ethylcyclopropane	2.48	2.14	2.93	3.16	4	1.40±0.59 (A1)

R-152 (1,2-Difluoroethane)	3.08	3.03	3.11	-	-	4.15±0.41 (A1)
R-143 (1,1,2-Trifluoroethane)	3.07	3.00	3.12	-	-	6.20±1.55 (A1)
R-245fa (1,1,1,3,3-Pentafluoro-propane)	3.08	3.05	3.09	-	-	7.70±0.77 (A1)

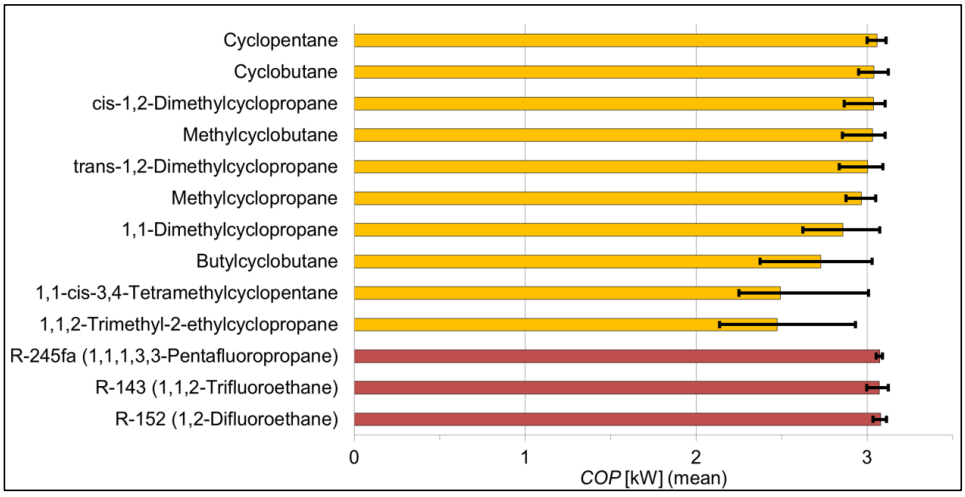


Figure 4. Ranking according to mean value of the *COP* of the best performing compounds. The 95%-confidence interval (thin black bars) was obtained from the uncertainty analysis with respect to the fluid properties, the identified cyclic hydrocarbons (yellow) are compared to commonly used refrigerants (red).

One of the identified cyclic fluids, cyclopentane, is also implemented in the well-established REFPROP 9.0 database [51]. Using the same process models and conditions, but using the Helmholtz-based EoS provided by REFPROP, the *COP* of cyclopentane was 3.05. This value lies inside the 95% confidence interval for cyclopentane calculated using PR-EoS (see Table 4) and corresponds to a relative deviation to the *COP* mean value of only 1%.

Considering the upper bound of the 95% confidence interval of the real fluids in Table 4, it can be seen that the *COP* values of the identified working fluids come close to the COP value of the optimal virtual fluids. This implies that the sampling based reverse engineering approach succeeded in identifying real working fluids which could provide a high *COP*. This identification has been achieved without evaluating all considered 2126 cyclic compounds in the cycle model and without solving a product design optimization problem. This adds credit to the effectiveness of the Monte Carlo sampling concept employed in the novel methodology for screening large chemical database. Furthermore, property uncertainty information is taken into account for the analysis of the performance of the identified fluids.

This study focused on cyclic hydrocarbons. Out of 2126 cyclic compounds only the property values of the 10 compounds in Table 4 are close to the top performing sample fluids. It is observed that compounds, which have had a distance function d_{xy}^{tot}>0.2, have been too far away from the virtual fluids to give a feasible solution. This boundary corresponds to an average relative difference between the property value of the virtual fluid and the real fluid of 20%. This value is specific for the given study. However, it confirms that the property search space is highly non-continuous, which makes standard (non-sampling based) numerical optimization tools [1] difficult or even impossible. Furthermore, the molecules identified by the algorithm are

comparatively small cyclic hydrocarbons. Large cyclic, aromatics or polycyclic compounds had in particular a critical temperature, which was much larger than the critical temperature of the optimal virtual fluids.

The *COP* uncertainty range of the identified compounds is overlapping with the 95%-confidence interval. The best performing compound with the smallest uncertainty range in *COP*, a low *GWP* (<150 [60]) and a high flammability limit (for safer operation) for the considered cycle is cyclopentane.

The upper bounds of the identified real fluids are very close to each other, whereas the lower bounds differ largely. This can be explained by the algorithm for the calculation of the distance value between the property of the real and the virtual fluid. The distance is the difference between the property value of the virtual fluid and the boundary of the uncertainty range of the real fluid. Hence, a statistically optimistic approach is taken for the identification of the working fluids. This optimistic approach is reflected in the uncertainty range of the model output (*COP*), where the upper bound shows the statistically best possible performance of the real fluid, which is based on the working fluid properties closest to the optimal virtual fluid. However, the uncertainty analysis after the reverse engineering algorithm considers the property uncertainty over the whole range from lower to upper bounds. In this way, it is revealed which fluids can statistically have a lower performance (conservative approach). Hence, the property based uncertainty analysis is an important complementary information after the identification of suitable working fluids using the sampling based reverse engineering approach.

The identified compounds have similar *COP* as commonly used refrigerants. However, cyclic compounds have a much lower impact on the climate in agreement to present regulations on working fluids. However, due to their high flammability safety-measures need to be considered, when the compounds are further investigated experimentally. In particular small rings (e.g. cyclopropane- and cyclobutane-compounds) often suffer from ring tensions and can be instable [70], which is also reflected in their small value for the lower flammability limit.

Table 5 includes the properties of the optimal virtual fluids. The identified properties of the best performing virtual fluids can also serve as target properties for further studies, e.g. the identification of optimal mixture compositions. Hence, the sampling base approach for reverse engineering is not limited to the example of pure component fluid design shown in this work.

Table 5. Fluid properties of the optimal virtual (sample) fluids.

Virtual fluid	T_c [K]	P_c [kPa]	ω [-]	MW [g/mol]	$C_p(T_{cond})$ [J/(molK)]	$C_p(T_{sink}+10)$ [J/(molK)]
1	572.06	5046.58	0.098	66.51	56.32	74.04
2	602.22	2976.62	0.065	130.98	46.87	56.71
3	580.78	7321.58	0.194	58.89	83.02	91.92
4	512,16	4203.05	0.4744	114.77	69.93	75.19

Table 6. Working fluid properties and uncertainties for two selected compounds: T_c is the critical temperature, P_c the critical pressure, ω the acentric factor, MW the molecular weight, $C_p(T_{cond})$ the heat capacity at the condensation temperature (lowest cycle temperature) and $C_p(T_{sink}+10)$ is the heat capacity at the heighest cycle temperature.

Rank	Working fluid name	T_c [K]	P_c [kPa]	ω [-]	MW [g/mol]	$C_p(T_{cond})$ [J/(molK)]	$C_p(T_{sink}+10)$ [J/(molK)]	closest to virtual fluid
1	Cyclopentane	511.74±0.18	4515.00±0.19	0.190±0.001	70	84.74±0.86	134.84±1.36	3
7	1,1-Dimethyl-cyclopropane	475±17	4190.00±5.49	0.117±0.046	70	108.01±1.10	153.98±1.56	1

3.2. Comparison of uncertainty propagation of cylopentane and 1,1-dimethylcyclopropane

Compared to the top performing cyclic hydrocarbon (based on mean value), cyclopentane, the uncertainty ranges are in particular much larger for the fluids ranked from 7 to 10 (e.g. 1,1-dimethylcyclopropane). The differences in the output uncertainty ranges originate in the uncertainties of the respective fluid properties (see Table 6). A high input uncertainty in a sensitive property can cause a high process output uncertainty. In particular the uncertainty of the acentric factor, which has been identified to be the most sensitive property in this study, is large for 1,1-dimethyl-cyclopropane compared to cyclopentane. Furthermore, also the uncertainties in the critical properties are by a magnitude higher for 1,1-Dimethyl-cyclopropane. Acentric factor and critical properties are the input parameters to the PR-EoS, which is used to calculate the enthalpies, entropies and fugacities of the fluid in the cycle. The uncertainty propagation of clopentane and 1,1-dimethylcyclopropane is further illustrated and analyzed in detail in the following.

The Monte Carlo simulation results of two compounds, cyclopentane and 1,1-dimethylcyclopropane, are shown here in more detail. The simulation results for each for the fluid property samples can be visualized. Figure 5 depicts the distribution of the *COP* simulation results, the temperature-heat (*T*-*Q*) diagram as well as the logarithmic-pressure-enthalpy log(*P*)-*h* diagram of the two compounds. Diagrams show an overlay plot of the simulation results. Hence, the uncertainty subject to the fluid properties is visualized through a varying band for both compounds. The black lines show the mean value performance. For cyclopentane the width of the output range is very narrow, whereas for 1,1-dimethylcyclopropane it is much larger. There is no variation in the heat sink line, since it has been fixed and defined as constraints.

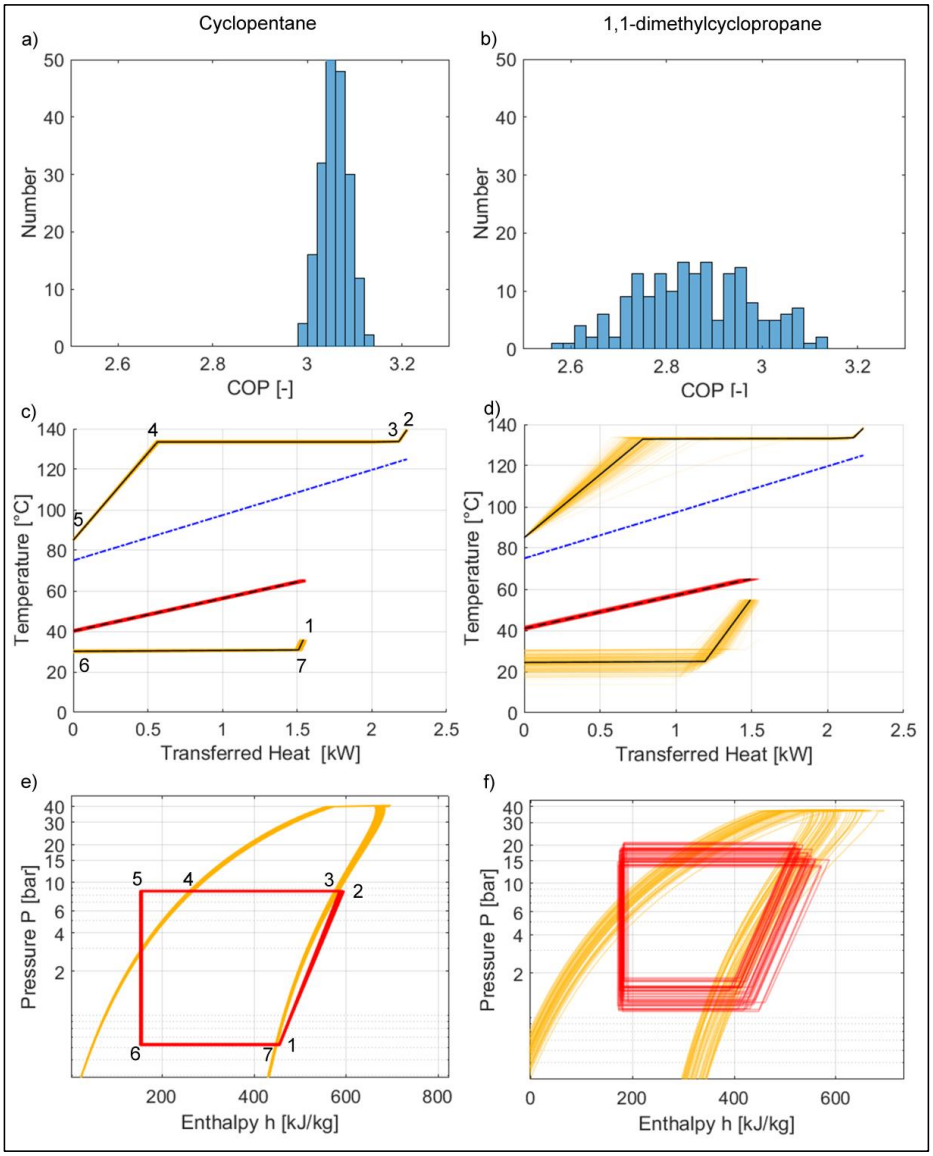


Figure 5. Representation of uncertainty with respect to the fluid properties in distribution of model output uncertainty of *COP*, *T*-*Q* diagram and log(*P*)-*h* diagram for cyclopentane and for 1,1-dimethylcyclopropane. The numbers refer to the states of the heat pump cycle according to Figure 1.

The visualization of the simulation results allows analysing, where in the cycle the property uncertainties lead to the largest process model uncertainty for a given compound. In the case of 1,1-dimethylcyclopropane, the compression and evaporation process shows the largest variation.

The distributions in Figure 5 illustrate the uncertainty range of *COP*. The property-related uncertainty of 1,1-dimethylcyclopropane is much larger than the one of cyclopentane (as it is also reflected in the 95% confidence intervals). Thus, the identification of cyclopentane as a top performing working fluid can be considered to be more reliable with respect to its fluid properties.

Even small variations of input property parameters can lead to large output uncertainty. This has already been shown for organic Rankine cycles by Frutiger et al. [45]. In this study, it can be confirmed for a heat pump system, that property uncertainties can vary significantly and *should not be neglected* for the selection of fluids.

4. Conclusion

The study presented a novel methodology to identify suitable working fluids for heat pump systems. A reverse engineering approach of the fluid selection based on Monte Carlo sampling of the fluid property search space was shown. Subsequently uncertainty propagation with respect to the fluid properties was performed. The methodology was applied for the selection of a pure component working fluid for an industrial heat pump system used for the recovery of heat from spray drying facilities in dairy industries. The study focused on the identification of suitable cyclic hydrocarbon based working fluids.

The following are the main conclusions from the systematic screening and the uncertainty analysis:

- The sampling based reverse engineering method identified top performing working fluids, but avoided solving computationally demanding molecular design problems.
- Real fluids could be identified based on the optimal target properties and taking into account the real fluid property uncertainties.
- The methodology combined an optimistic approach with respect to uncertainties (distance function to uncertainty bound) with a conservative approach (subsequent property uncertainty propagation through cycle). Thereby, the methodology gave an additional dimension to the fluid selection process.
- The considered case study focused on the identification of cyclic hydrocarbons. For the given application small cyclic hydrocarbons showed the best performance, although the output uncertainty range varies largely. Cyclopentane obtained the largest *COP* and the smallest uncertainty range. Furthermore, the calculation of the lower flammability limit of the compounds showed that safety-measures are needed, if the cyclic hydrocarbons are further investigated in experiments.

We believe that the novel reverse engineering approach can be useful for process developers of thermodynamic cycles, because it allows a simple identification of working fluids through

the application of Monte Carlo methods. Furthermore, the methodology has been formulated in a generic way and it is possible to apply it to product-process design problems in process engineering beyond working fluids for thermodynamic cycles.

References

[1] Gani R, Zhang L, Kalakul S, Cignitti S. Computer-Aided Molecular Design and Property Prediction. vol. 39. Elsevier; 2017.

[2] Harper PM, Gani R, Kolar P, Ishikawa T. Computer-aided molecular design with combined molecular modeling and group contribution. Fluid Phase Equilib 1999;158–160:337–47.

[3] Gani R, Ng KM. Product design - Molecules, devices, functional products, and formulated products. Comput Chem Eng 2015;81:70–9.

[4] Boone K, Abedin F, Anwar MR, Camarda KV. Molecular Design in the Pharmaceutical Industries. vol. 39. Elsevier; 2017.

[5] Linke P, Papadopoulos A, Seferlis P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. Energies 2015;8:4755–801.

[6] Cussler EL, Moggridge GD. Introduction to chemical product design. Chem. Prod. Des., 2011, p. 1–15.

[7] Frutiger J, Marcarie C, Abildskov J, Sin G. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models - an application to the heat of combustion. J Chem Eng Data 2016;61:602–13.

[8] Frenkel M, Chirico RD, Diky V, Yan X, Dong Q, Muzny C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. J Chem Inf Model 2005;45:816–38.

[9] Diky V, Chirico RD, Muzny CD, Kazakov AF, Kroenlein K, Magee JW, et al. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 8. Properties of Material Streams and Solvent Design. Chem Inf Model 2013;53:249–66. doi:10.1021/ci300470t.

[10] Rowley RL, Wilding W V., Oscarson, L. J, Knotts TA, Giles NF. DIPPR® Data Compilation of Pure Chemical Properties, Design Institute for Physical Properties, AIChE, New York, NY 2014.

[11] Dong Q, Chirico RD, Yan X, Hong X, Frenkel M. Uncertainty Reporting for Experimental Thermodynamic Properties. J Chem Eng Data 2005;50:546–50.

[12] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilib 2012;321:25–43.

[13] Adjiman CS, Androulakis IP, Floudas CA. Global optimization of MINLP problems in process synthesis and design. Comput Chem Eng 1997;21:S445–50. doi:10.1016/S0098-1354(97)87542-4.

[14] Cignitti S, Zhang L, Gani R. Computer-aided Framework for Design of Pure, Mixed and Blended Products. 12th Int Symp Process Syst Eng 25th Eur Symp Comput Aided Process Eng 2015;37:2093–8.

[15] Frutiger J, Abildskov J, Sin G. Global sensitivity analysis of computer-aided molecular design problem for the development of novel working fluids for power cycles. 26th Eur. Symp. Comput. Aided Process Eng., 2016, p. 283–8.

[16] Roskosch D, Atakan B. Reverse engineering of fluid selection for thermodynamic cycles with cubic equations of state, using a compression heat pump as example. Energy 2015;81:202–12. doi:10.1016/j.energy.2014.12.025.

[17] Gani R. Chemical product design: challenges and opportunities. Comput Chem Eng 2004;28:2441–57. doi:10.1016/j.compchemeng.2004.08.010.

[18] Mikielewicz D, Mikielewicz J. Criteria for selection of working fluid in low-temperature ORC.

Chem Process Eng Chem I Proces 2016;37:429–40.
[19] Ng LY, Chemmangattuvalappil NG, Dev VA, Eden MR. Chapter 1 – Mathematical Principles of Chemical Product Design and Strategies. vol. 39. 2017. doi:10.1016/B978-0-444-63683-6.00001-0.
[20] Maranas C. Optimal molecular design under property prediction uncertainty. AICHE J 1997;43:1250–64.
[21] Diwekar UM, Xu W. Improved genetic algorithms for deterministic optimization and optimization under uncertainty. Part I. Algorithms development. Ind Eng Chem Res 2005;44:7132–7.
[22] Ng LY, Chemmangattuvalappil NG, Ng DKS. Robust chemical product design via fuzzy optimisation approach. Comput Aided Chem Eng 2014;34:387–92.
[23] Ng LY, Chemmangattuvalappil NG, Ng DKS. Optimal Chemical Product Design via Fuzzy Optimisation based Inverse Design Techniques. vol. 33. Elsevier; 2014. doi:10.1016/B978-0-444-63456-6.50055-7.
[24] Bühler F, Fridolin M, Huang B, Andreasen JG, Elmegaard B. Mapping of low temperature heat sources in Denmark. ECOS 2015 28th Int Conf Effic Cost, Optim Simul Environ Impact Energy Syst 2015.
[25] Cignitti S, Frutiger J, Zühlsdorf B, Bühler F, Andreasen JG, Müller F, et al. Forbedring af industrielle processers energieffektivitet. Dansk Kemi 2016;97.
[26] Bao J, Zhao L. A review of working fluid and expander selections for organic Rankine cycle. Renew Sustain Energy Rev 2013;24:325–42.
[27] Brown JS, Brignoli R, Daubman S. Methodology for estimating thermodynamic parameters and performance of working fluids for organic Rankine cycles. Energy 2014;73:818–28.
[28] Brown JS, Brignoli R, Quine T. Parametric investigation of working fluids for organic Rankine cycle applications. Appl Therm Eng 2015;90:64–74.
[29] Stijepovic MZ, Linke P, Papadopoulos AI, Grujic AS. On the role of working fluid properties in Organic Rankine Cycle performance. Appl Therm Eng 2012;36:406–13.
[30] Stijepovic MZ, Papadopoulos AI, Linke P, Stijepovic V, Grujic AS, Kijevčanin M, et al. Organic Rankine Cycle system performance targeting and design for multiple heat sources with simultaneous working fluid selection. J Clean Prod 2017;142:1950–70.
[31] Zhai H, Shi L, An Q. Influence of working fluid properties on system performance and screen evaluation indicators for geothermal ORC (organic Rankine cycle) system. Energy 2014;74:2–11.
[32] Desideri A, Gusev S, Broek M Van Den, Lemort V, Quoilin S. Experimental comparison of organic fluids for low temperature ORC systems for waste heat recovery applications 2015;97:460–9.
[33] Hærvig J, Sørensen K, Condra TJ. Guidelines for optimal selection of working fluid for an organic Rankine cycle in relation to waste heat recovery. Energy 2016;96:592–602.
[34] Xu J, Yu C. Critical temperature criterion for selection of working fluids for subcritical pressure Organic Rankine cycles. Energy 2014;74:719–33.
[35] Papadopoulos AI, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward Optimum Working Fluid Mixtures for Organic Rankine Cycles using Molecular Design and Sensitivity Analysis. Ind Eng Chem Res 2013;52:12116–33.
[36] Mavrou P, Papadopoulos AI, Seferlis P, Linke P, Voutetakis S. Selection of working fluid mixtures for flexible Organic Rankine Cycles under operating variability through a systematic nonlinear sensitivity analysis approach. Appl Therm Eng 2015;89:1054–67. doi:10.1016/j.applthermaleng.2015.06.017.
[37] Kucherenko S, Rodriguez-Fernandez M, Pantelides C, Shah N. Monte Carlo evaluation of derivative-based global sensitivity measures. Reliab Eng Syst Saf 2009;94:1135–48.
[38] Papadopoulos AI, Stijepovic M, Linke P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. Appl Therm Eng 2010;30:760–9.
[39] Papadopoulos AI, Linke P, Seferlis P. Integrated Multiobjective Molecular and Process Design: Operational and Computational Frontiers. vol. 39. Elsevier; 2017.

[40] Palma-Flores O, Flores-Tlacuahuac A, Canseco-Melchor G. Optimal molecular design of working fluids for sustainable low-temperature energy recovery. Comput Chem Eng 2015;72:334–49.
[41] Cignitti S, Andreasen JG, Mansouri SS, Haglind F, Woodley JM, Abildskov J. Integrated working fluid-cycle design of an industrial organic Rankine cycle unit for low temperature waste heat recovery. Appl Energy, Submitt 2017.
[42] Molina-Thierry DP, Flores-Tlacuahuac A. Simultaneous optimal design of organic mixtures and rankine cycles for low-temperature energy recovery. Ind Eng Chem Res 2015;54:3367–83.
[43] Lampe M, Stavrou M, Schilling J, Sauer E, Gross J, Bardow A. Computer-aided Molecular Design in the Continuous-Molecular Targeting Framework using Group-Contribution PC-SAFT. Comput Chem Eng 2015. doi:10.1016/j.compchemeng.2015.04.008.
[44] Schilling J, Lampe M, Gross J, Bardow A. One-stage approach for the integrated design of ORC processes and working fluid using PC-SAFT. Comput Aided Chem Eng 2016;38:1335–40. doi:10.1016/B978-0-444-63428-3.50227-7.
[45] Frutiger J, Andreasen JG, Liu W, Spliethoff H, Haglind F, Abildskov J, et al. Working fluid selection for organic Rankine cycles - impact of uncertainty of fluid properties. Energy 2016;109:987–97.
[46] Frutiger J, Bell I, O'Connell JP, Kroenlein K, Abildskov J, Sin G. Uncertainty assessment of equations of state with application to an arganic rankine cycle. Mol Phys 2016.
[47] Zühlsdorf B, Bühler F, Mancini R, Cignitti S, Brian Elmegaard. High Temperature Heat Pump Integration using Zeotropic Working Fluids for Spray Drying Facilities. 12th IEA Heat Pump Conf 2017.
[48] Bühler F, Nguyen T Van, Elmegaard B. Energy and exergy analyses of the Danish industry sector. Appl Energy 2016;184:1447–59. doi:10.1016/j.apenergy.2016.02.072.
[49] Peng D-Y, Robinson DB. A New Two-Constant Equation of State. Ind Eng Chem Fundam 1976;15:59–64.
[50] Papadopoulos AI, Stijepovic M, Linke P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. Appl Therm Eng 2010;30:760–9.
[51] Lemmon E, Huber M, McLinden M. Reference fluid thermodynamic and transport properties-REFPROP, standard reference database 23, version 8.0, National Institute of Standard and Technology; 2007.
[52] Bell I, Wronski J, Quoilin S, Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. Ind Eng Chem Res 2014;53:2498–508.
[53] Kelley C. T. Solving nonlinear equations with Newton's method. 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics; 2003.
[54] Odele O, Macchietto S. Computer aided molecular design: a novel method for optimal solvent selection. Fluid Phase Equilib 1993;82:47–54.
[55] Aly F, Lee L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. Fluid Phase Equilib 1981;6:169–79.
[56] Seber G, Wild C. Nonlinear Regression. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1989.
[57] Sobol IM. On quasi-Monte Carlo integrations. Math Comput Simul 1998;47:103–12. doi:10.1016/S0378-4754(98)00096-2.
[58] Halton JH. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer Math 1960;2:84–90. doi:10.1007/BF01386213.
[59] Sobol' IM, Kucherenko S. Derivative based global sensitivity measures and their link with global sensitivity indices. Math Comput Simul 2009;79:3009–17. doi:10.1016/j.matcom.2009.01.023.
[60] European Parliament and council. Regulation (EC) No 1005/2009 of the European Parliament and of the council. Off J Eur Communities 2000;L 269:1–15.
[61] Hwang Y, Ohadi M, Radermacher R. Natural refrigerants. Mech Eng 1998;120:96–9.
[62] Hukkerikar AS. Development of pure component property models for chemical product-process design and analysis. 2013.

[63] Joback K, Reid R. Estimation of pure-component properties from group-contribution. Chem Eng Commun 1987;57:233–43.
[64] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.
[65] Iman R, Conover W. A distribution-free approach to inducing rank correlation among input variables. Commun Stat Part B-Simulation Comput 1982;11:311–34.
[66] Frutiger J, Marcarie C, Abildskov J, Sin G. Group-Contribution based Property Estimation and Uncertainty analysis for Flammability-related Properties. J Hazard Mater 2016;318:783–93. doi:10.1016/j.jhazmat.2016.06.018.
[67] Hukkerikar AS, Kalakul S, Sarup B, Young DM, Sin G, Gani R. Estimation of environment-related properties of chemicals for design of sustainable processes: development of group-contribution+ (GC+) property models and uncertainty analysis. J Chem Inf Model 2012;52:2823–39.
[68] Forster P, Ramaswamy V. Changes in atmospheric constituents and in radiative forcing, chapter 2. In: Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt MT and HLM, editor. Clim. Chang. 2007 Phys. Sci. Basis. Contrib. Work. Gr. I to Fourth Assess. Rep. Intergov. Panel Clim. Chang., Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2007, p. 129–234. doi:10.1103/PhysRevB.77.220407.
[69] Wilson DP, Kujak S, Leary JMO, Kennoy DH, Kusmierz A, Patnaik V, et al. ANSI/ASHRAE Standard 34 Designation and Safety Classification of Refrigerants. vol. 34. 2010.
[70] March J. Advanced organic chemistry: reactions, mechanisms, and structure. New York: John Wiley & Sons, Inc.; 1992.

Paper H

Global sensitivity analysis of computer-aided molecular design problem for the development of novel working fluids for power cycles

Jérôme Frutiger, Jens Abildskov, Gürkan Sin

Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

published in Proceedings of the 26th European Symposium on Computer Aided Process Engineering *ESCAPE* 26 (ISBN: 978-0-444-63428-3), 38, pages: 283–288, 2016,
DOI: http://dx.doi.org/10.1016/B978-0-444-63428-3.50052-7

and presented at *ESCAPE* 26, June 12-15 2016, Portorož, Slovenia.

Abstract

This study compares two methods for global sensitivity analysis as a new approach for the identification and ranking of target properties in molecular design problems: A modified Morris Screening technique and Monte Carlo based standard regression. The two methodologies are highlighted in a case study involving the design of a working fluid for an Organic Ranking Cycle (ORC) design for power generation. Morris Screening is found to be favorable over Monte Carlo based standard regression. Monte Carlo based standard regression cannot be applied, because the current model cannot be sufficiently linearized. For Morris Screening techniques the critical temperature, the critical pressure and the acentric factor of the working fluid has been identified as the target properties with the highest sensitivity to the net power output of the cycle.

1. Introduction

Power cycles, like Organic Rankine Cycle (ORC), allow converting industrial waste heat into usable electrical energy. In order to optimize the heat transfer process and the power generation, the influence of the working fluid is crucial. Multi-criteria database search and Computer Aided Molecular Design (CAMD) can be applied to generate, test and evaluate promising pure component/mixture candidate as process fluids to help optimize cycle design and performance [1].

The first step in the molecular design problem formulation is commonly the identification of target properties. In many molecular design applications the expert knowledge or literature studies is used as a source for target property identification [2]. We propose a new approach for the systematic analysis of target properties of molecular design problems with respect to the model output: the usage of sensitivity analysis as a global tool to address major target property identification.

In this study, we compare two methods for global sensitivity analysis, to identify and rank relevant target properties of working fluids: 1) Morris Screening techniques 2) Monte Carlo based standard regression (SRC) [3]. The methodologies are highlighted in a case study involving an ORC design for energy recovery from low-heat waste streams. The two models are both well-known as an efficient way of performing global sensitivity analysis. The advantage of Morris's method is that it does not rely on restricted assumptions (e.g. linearizable model), but it does not include parameter correlation. On the other hand, SRC takes into account parameter interdependency and is based on well-established regression principles, but it is necessary to test, if the model fulfils the criterion of being linearizable. The goal of this study is to compare the performance of these methods in the context of a molecular design problem.

2. Methodology

2.2. Process and property model

In this study the CAMD problem for the development of novel working fluids for organic Rankine cycles (ORC) is formulated as a mathematical optimization problem. It integrates both an ORC system model and thermodynamic property models.

The detailed process model with the corresponding process specifications can be found in the work of Andreasen et al. (2014). The ORC process is sketched in Figure 1.

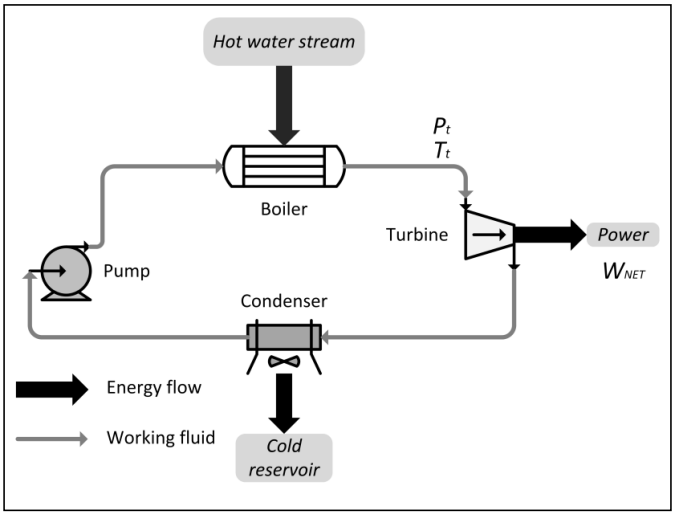


Figure 1. Scheme of the ORC process.

The cycle consists of a pump, a boiler, a turbine and a condenser. The working fluid circulates around the cycle. It is pressurized as a liquid by the pump, takes up heat from a hot water stream through the boiler system and is evaporated. Then the hot pressurized vapor expands in the turbine, producing mechanical power (see Figure 1). The ORC objective function is the net power output W_{NET}, which is the difference between the power production from the turbine and the power consumption of the pump.

Peng-Robinson equation of state (PR EoS) is used for the estimation of the thermodynamic properties. The property model is implemented following the work of Liu et al. (2014). PR EoS is selected, because of its relatively small number of required fluid parameters, which makes it suitable for molecular design problems: Critical temperature T_c, critical pressure P_c, acentric factor ω, molecular weight MW and the correlation constants (A, B, C, D, E) for the temperature-dependent heat capacity C_p defined by Aly and Lee (1981). The molecular design task is to identify the best fluid parameter set (T_c, P_c, ω, MW, A, B, C, D, E) giving the highest net power output W_{NET}. A given optimal set of properties can not necessarily be realized by a real fluid. However, the best fluid parameter set provides the desired set of values that maximizes W_{NET}.

2.2. Global sensitivity analysis with respect to fluid property parameters

In this study we focus exclusively on a global sensitivity analysis of the fluid parameters using Morris Screening techniques and Monte Carlo based Standardized Regression Coefficients (SRC). The procedures follow the work of Sin et al. (2009). The range of the fluid parameters (design space) is specified by selecting a class of chemicals, i.e. hydrocarbons, alcohols, halogen-containing compounds, and taking the lowest parameter value as the lower bound of the parameter range and the respective highest parameter value as the higher bound of the design space. The samples for the sensitivity analysis are taken from this design space.

Modified Morris Screening

The method tries to estimate the distribution of the elementary effects (EE_j) of each fluid parameter θ_j on the model output W_{NET}. This distribution function of EE_j, denoted F_j, represents the distribution of the effects of fluid parameter j on the output. EE_j is obtained from

$$EE_j = \frac{W_{NET}(\theta_1, \theta_1, ..., \theta_j, ..., \theta_M) - W_{NET}(\theta_1, \theta_1, ..., \theta_j + \Delta, ..., \theta_M)}{\Delta} \quad (1)$$

Here Δ is a perturbation factor of θ_j, $W_{NET}(\theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_M)$ is model output evaluated at fluid parameters $(\theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_M)$ and $W_{NET}(\theta_1, \theta_2, \ldots, \theta_j+\Delta, \ldots, \theta_M)$ is the model output corresponding to a change Δ in θ_j. M is the number of parameters. The overall picture of elementary effects can be analyzed looking at the mean and standard deviation of the distribution function F_j. F_j is estimated by repeating the model evaluations of the elementary effects, EE_j, at randomly sampled points in the fluid parameter input space. This is repeated a number of times r. Morris (1991), suggested an effective one-factor-at-a-time (OAT) design, where the calculation of one elementary effect for each input requires (M+1) model simulations, giving a total number of simulations of r (M+1). Each input parameter, θ_j can only take values corresponding to a predefined grid in which the range of each parameter is subdivided into p levels. Usually p can take the values of 4, 6 or 8 and r is between 4 and 15. However, in our case we chose a grid containing p=64 levels with r=120 repetitions. This modified Morris Screening allows having a larger coverage of the parameter space.

Monte Carlo based standard regression

The first step of the Monte Carlo based approach is sampling from the defined fluid parameter design space. The Latin Hypercube Sampling method [8] was used for probabilistic sampling of the fluid parameter space. 400 samples were selected where each sample consists of a parameter set. Correlation between the input parameters is taken into account using the rank-based method of Iman and Conover (1982). The sampled input parameter sets are propagated through the model by performing one simulation for each input parameter sample. By constructing linear regression models on the outputs obtained from the Monte Carlo procedure, it is possible to obtain the standardized regression coefficient (SRC) β_j of parameter set θ_j given by

$$\frac{W_{NET} - \mu_{WNET}}{\sigma_{WNET}} = \sum_{j=1}^{M} \beta_j \cdot \frac{\theta_j - \mu_{\theta j}}{\sigma_{\theta j}} + \varepsilon_j \quad (2)$$

W_{NET} is model output, ε_j is the error of the regression model, $\mu_{\theta j}/\sigma_{\theta j}$ and μ_{NET}/σ_{NET} are the corresponding mean and standard deviations of the parameters and the model output. The SRC is a measure of how strongly the parameter contributes to the net power output

3. Results and discussion

3.1. Morris Screening

Figure 2 shows the results of Morris Screening technique by comparing the mean and the standard deviations of the distribution function, F_j, of the elementary effects of each fluid parameter. Furthermore, there are two lines depicted in Figure 2, which correspond to mean ± standard deviation/$\sqrt{r}$. The two lines help interpret the effects of the fluid parameters on the outputs. If a fluid parameter lies inside the cone, its effect on the model output is negligible.

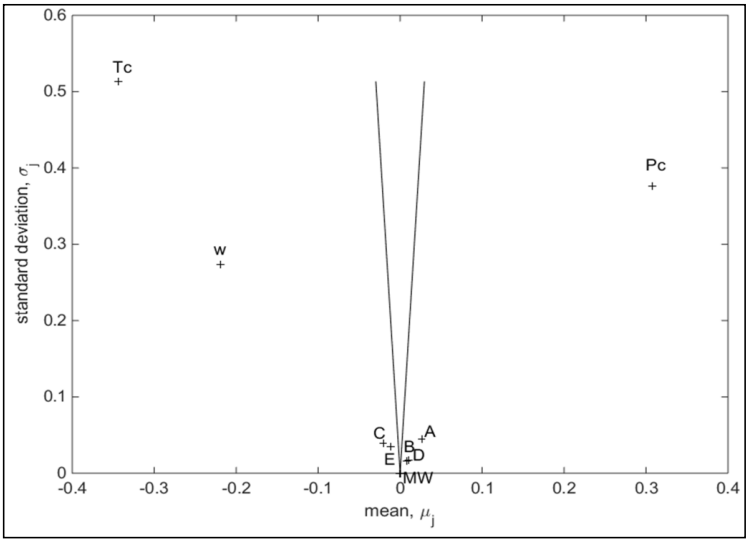


Figure 2. Estimated means and standard deviations of the distribution function of the scaled elementary effects of the parameters.

None of the fluid parameters lie inside the cone. Hence, none of these are insignificant. T_c, P_c and ω have a high standard deviation. This indicates that these are involved in high non-linear interactions. This makes sense, because these parameters appear in the PR EoS, which is highly non-linear. *MW* has been found to have a zero standard deviation with non-zero mean, which means that *MW* has a linear effect on the output. The heat capacity constants *A*, *B*, *C*, *D* and *E* have small standard deviations and mean values. By computing the absolute mean of the distribution function, F_j, one can rank the parameters according to their significance – the higher the mean value, the more significant the effect of the fluid parameter on the net power output (see Table 1).

Table 1. Estimated absolute means of the distribution of elementary effects.

Rank	θ_i	Iμ_iI	Rank	θ_i	Iμ_iI
1	T_c	0.34	6	*E*	0.011
2	P_c	0.31	7	*D*	0.010
3	ω	0.22	8	*B*	0.0077
4	*A*	0.027	9	*MW*	0
5	*C*	0.021			

As it can be seen in Table 1, the net power output is highly sensitive to the PR-EOS input parameters T_c, P_c and ω, whereas the heat capacity constants (*A*, *B*, *C*, *D* and *E*) and the molecular weight *MW* have lower significance for the overall product-process model. The major influence of critical properties (in this case T_c and P_c) and molecular complexity (here

ω) of the working fluid on the net power out has also been reported by Invernizzi (2013). The reason is that these parameters determine the position and the shape of the saturation limit and therefore, in which temperature and pressure region of the fluid state, the cycle will operate. Morris Screening allows us to quantitatively show that the PR-EOS parameters (T_c, P_c and ω) have a higher impact on the net power output than the heat capacity constants and molecular weight. Hence, for the fluid design problem it is of higher priority to match desired values for T_c, P_c and ω than MW and A, B, C, D, E. Furthermore, the method is not based on the assumption that the model is linearizable. However, its disadvantage is that it does not take into account correlation or interdependence of the fluid parameters, as opposed to the Monte Carlo based methods. Therefore, it has been necessary to filter out unfeasible solutions.

3.2. Monte Carlo based standard regression

The regression coefficients – SRCs – of the fluid parameters were obtained from linear least square regression using Eq. 2.

Table 2. Standardized Regression Coefficients (SRCs)

Rank	θ_i	ISRC_iI	Rank	θ_i	ISRC_iI
1	T_c	0.58	6	ω	0.15
2	C	0.33	7	P_c	014
3	E	0.30	8	A	0.11
4	B	0.20	9	D	0.09
5	MW	0.19	$\boldsymbol{R^2}$	**0.26**	

The SRCs were ranked according to their absolute value (see Table 2). According to the analysis the critical temperature T_c has by far the highest influence compared to the other parameters. This is in agreement with the analysis from Morris Screening technique. However, the heat capacity constants have a higher *SRC* than the acentric factor ω and the critical pressure P_c. The Pearson correlation coefficient R_2 (see Table 2) indicates that the degree of linearization is low. The model cannot be sufficiently linearized in order to fulfill the assumption of the Monte Carlo standard regression theory. In that sense the standard regression coefficients obtained should be considered with caution and as less reliable than the results obtained from Morris Screening.

4. Conclusion

Two methods were investigated to perform sensitivity analysis for molecular design problems and whether they can be applied for the molecular design problem of an ORC working fluid. Modified Morris Screening is found to be favorable over Monte Carlo based standard regression. The Morris Screening technique identifies the critical temperature T_c, the critical pressure P_c and the acentric factor ω as most influential target properties of the working fluid. Hence, in a molecular design problem formulation, these target properties have a higher importance on the model output than the others. We suggest that more emphasis should be given on the measurement and prediction of the identified target properties.

References

[1] Linke P, Papadopoulos A, Seferlis P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. Energies 2015;8:4755–801.

[2] Gani R. Computer-aided methods and tools for chemical product design 2004;82:1494–504.

[3] Sin G, Gernaey K V., Eliasson Lantz A. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. Biotechnol Prog 2009;25:1043–53.

[4] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic Rankine cycles. Energy 2014;73:204–13.

[5] Liu W, Meinel D, Wieland C, Spliethoff H. Investigation of hydrofluoroolefins as potential working fluids in organic Rankine cycle for geothermal power generation. Energy 2014;67:106–16.

[6] Aly F, Lee L. Self-consistent equations for calculating heat capacity, enthalpy, and entropy the ideal gas. Fluid Phase Equilib 1981;6:169–79.

[7] Morris MD. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 1991;33:161–74.

[8] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.

[9] Iman R, Conover W. A distribution-free approach to inducing rank correlation among input variables. Commun Stat Part B-Simulation Comput 1982;11:311–34.

[10] Invernizzi CM. The Organic Rankine Cycle. Closed Power Cycles, Thermodyn. Fundam. Appl., London: Springer Verlang; 2013, p. 117–76.

Technical University of Denmark
Anker Engelundsvej 1
Building 101A
DK-2800, Kgs. Lyngby
Denmark
CVR-nr. 30 06 09 46
Phone: (+45) 45 25 25 25
Email: dtu@dtu.dk
www.dtu.dk